EXAMPLE 3

Cloning of the genes expressing 17-30 kDa antigens from ST-CF

Isolation of CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5% (w/v) and 5% (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with an 8 M urea buffer containing 0.5% (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a Prep Cell (BioRad) in a matrix of 16% polyacrylamide under an electrical gradient. Fractions containing pure proteins with an molecular mass from 17-30 kDa were collected.

Isolation of CFP22

Anti-CFP29, reacting with CFP29 was generated by immunization of BALB/c mice with crushed gel pieces in RIBI adjuvant (first and second immunization) or aluminium hydroxide (third immunization and boosting) with two week intervals. SDS-PAGE gel pieces containing 2-5 µg of CFP29 were used for each immunization. Mice were boosted with antigen 3 days before removal of the spleen. Generation of a monoclonal cell line producing antibodies against CFP29 was obtained essentially as described by Köhler and Milstein (1975). Screening of

supernatants from growing clones was carried out by immunoblotting of nitrocellulose strips containing ST-CF separated by SDS-PAGE. Each strip contained approximately 50 µg of ST-CF. The antibody class of anti-CFP29 was identified as IgM by the mouse monoclonal antibody isotyping kit, RPN29 (Amersham) according to the manufacturer's instructions.

CFP29 was purified by the following method: ST-CF was concentrated 10 fold by ultrafiltration, and ammonium sulphate precipitation in the 45 to 55% saturation range was performed. The pellet was redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography (Porath *et al.*, 1985) on an Affi-T gel column (Kem-En-Tec). Protein was eluted by a linear 1.5 to 0 M gradient of ammonium sulphate and fractions collected in the range 0.44 to 0.31 M ammonium sulphate were identified as CFP29 containing fractions in Western blot experiments with mAb Anti-CFP29. These fractions were pooled and anion exchange chromatography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5 and the elution was performed with a linear gradient from 0 to 500 mM NaCl. From 400 to 500 mM sodium chloride, rather pure CFP29 was eluted. As a final purification step the Mono Q fractions containing CFP29 were loaded on a 12.5% SDS-PAGE gel and pure CFP29 was obtained by the multi-elution technique (Andersen and Heron, 1993).

N-terminal sequencing and amino acid analysis

CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28 were washed with water on a Centricon concentrator (Amicon) with cutoff at 10 kDa and then applied to a ProSpin concentrator (Applied Biosystems) where the proteins were collected on a PVDF membrane. The membrane was washed 5 times with 20% methanol before sequencing on a Procise sequencer (Applied Biosystems).

CFP29 containing fractions were blotted to PVDF membrane after tricine SDS-PAGE (Ploug et al., 1989). The relevant bands were excised and subjected to amino acid analysis (Barkholt and Jensen, 1989) and N-terminal sequence analysis on a Procise sequencer (Applied Biosystems).

The following N-terminal sequences were obtained:

For CFP17: A/S E L D A P A Q A G T E X A V (SEQ ID NO: 17)
For CFP20: A Q I T L R G N A I N T V G E (SEQ ID NO: 18)
For CFP21: D P X S D I A V V F A R G T N (SEQ ID NO: 19)
For CFP22: T N S P L A T A T A T L H T W (SEQ ID NO: 20)
For CFP25: A X F D A E V V F A R G R F E (SEQ ID NO: 21)
For CFP28: X I/V Q K S L E L I V/T V/F T A D/Q E (SEQ ID NO: 22)
For CFP29: M N N L Y E D L A P V T E A A W A E I (SEQ ID NO: 23)

"X" denotes an amino acid which could not be determined by the sequencing method used, whereas a "/" between two amino acids denotes that the sequencing method could not determine which of the two amino acids is the one actually present.

Cloning the gene encoding CFP29

The N-terminal sequence of CFP29 was used for a homology search in the EMBL database using the TFASTA program of the Genetics Computer Group sequence analysis software package. The search identified a protein, Linocin M18, from *Brevibacterium linens* that shares 74% identity with the 19 N-terminal amino acids of CFP29.

Based on this identity between the N-terminal sequence of CFP29 and the sequence of the Linocin M18 protein from *Brevibacterium linens*, a set of degenerated primers were constructed for PCR cloning of the *M. tuberculosis* gene encoding CFP29. PCR reactions were containing 10 ng of *M. tuberculosis* chromosomal DNA in 1 × low salt Taq+ buffer from Stratagene supplemented with 250 µM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 µl reaction volume. Reactions

were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 15 sec., 55°C for 15 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

An approx. 300 bp fragment was obtained using primers with the sequences:

1: 5'-CCCGGCTCGAGAACCTSTACCGCGACCTSGCSCC (SEQ ID NO: 24)
2: 5'-GGGCCGGATCCGASGCSGCGTCCTTSACSGGYTGCCA (SEQ ID NO: 25)
-where S = G/C and Y = T/C

The fragment was excised from a 1% agarose gel, purified by Spin-X spinn columns (Costar), cloned into pBluescript SK II+ - T vector (Stratagene) and finally sequenced with the Sequenase kit from United States Biochemical.

The first 150 bp of this sequence was used for a homology search using the Blast program of the Sanger *Mycobacterium tuberculosis* database:

(http//www.sanger.ac.uk/projects/M-tuberculosis/blast_server).

This program identified a *Mycobacterium tuberculosis* sequence on cosmid cy444 in the database that is nearly 100% identical to the 150 bp sequence of the CFP29 protein. The sequence is contained within a 795 bp open reading frame of which the 5' end translates into a sequence that is 100% identical to the N-terminally sequenced 19 amino acids of the purified CFP29 protein.

Finally, the 795 bp open reading frame was PCR cloned under the same PCR conditions as described above using the primers:

3: 5'-GGAAGCCCCATATGAACAATCTCTACCG (SEQ ID NO: 26)

4: 5'-CGCGCTCAGCCCTTAGTGACTGAGCGCGACCG (SEQ ID NO: 27)

The resulting DNA fragments were purified from agarose gels as described above sequenced with primer 3 and 4 in addition to the following primers:

5: 5'-GGACGTTCAAGCGACACATCGCCG-3' (SEQ ID NO: 115)
6: 5'-CAGCACGAACGCGCCGTCGATGGC-3' (SEQ ID NO: 116)

Three independent cloned were sequenced. All three clones were in 100% agreement with the sequence on cosmid cy444.

All other DNA manipulations were done according to Maniatis *et al.* (1989).

All enzymes other than Taq polymerase were from New England Biolabs.

Homology searches in the Sanger database

For CFP17, CFP20, CFP21, CFP22, CFP25, and CFP28 the N-terminal amino acid sequence from each of the proteins were used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* database:

http://www.sanger.ac.uk/pathogens/TB-blast-server.html.

For CFP29 the first 150 bp of the DNA sequence was used for the search. Furthermore, the EMBL database was searched for proteins with homology to CFP29.

Thereby, the following information were obtained:

CFP17

Of the 14 determined amino acids in CFP17 a 93% identical sequence was found with MTCY1A11.16c. The difference between the two sequences is in the first amino acid: It is an A or an S in the N-terminal determined sequenced and a S in

MTCY1A11. From the N-terminal sequencing it was not possible to determine amino acid number 13.

Within the open reading frame the translated protein is 162 amino acids long. The N-terminal of the protein purified from culture filtrate starts at amino acid 31 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 132 amino acids, which corresponds to a theoretical molecular mass of 13833 Da and a theoretical pI of 4.4. The observed mass in SDS-PAGE is 17 kDa.

CFP20

A sequence 100% identical to the 15 determined amino acids of CFP20 was found on the translated cosmid cscy09F9. A stop codon is found at amino acid 166 from the amino acid M at position 1. This gives a predicted length of 165 amino acids, which corresponds to a theoretical molecular mass of 16897 Da and a pI of 4.2. The observed molecular weight in a SDS-PAGE is 20 kDa.

Searching the GenEMBL database using the TFASTA algorithm (Pearson and Lipman, 1988) revealed a number of proteins with homology to the predicted 164 amino acids long translated protein.

The highest homology, 51.5% identity in a 163 amino acid overlap, was found to a Haemophilus influenza Rd toxR reg. (HIHI0751).

CFP21

A sequence 100% identical to the 14 determined amino acids of CFP21 was found at MTCY39. From the N-terminal sequencing it was not possible to determine amino acid number 3; this amino acid is a C in MTCY39. The amino acid C can not be detected

on a Sequencer which is probably the explanation of this difference.

Within the open reading frame the translated protein is 217 amino acids long. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 33 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 185 amino acids, which corresponds to a theoretical molecular weigh at 18657 Da, and a theoretical pI at 4,6. The observed weight in a SDS-PAGE is 21 kDa.

In a 193 amino acids overlap the protein has 32,6% identity to a cutinase precursor with a length of 209 amino acids (CUTI_ALTBR P41744).

A comparison of the 14 N-terminal determined amino acids with the translated region (RD2) deleted in *M. bovis* BCG revealed a 100% identical sequence (mb3484) (Mahairas *et al.* (1996)).

CFP22

A sequence 100% identical to the 15 determined amino acids of CFP22 was found at MTCY10H4. Within the open reading frame the translated protein is 182 amino acids long. The N-terminal sequence of the protein purified from culture filtrate starts at amino acid 8 and therefore the length of the protein occurring in *M. tuberculosis* culture filtrate is 175 amino acids. This gives a theoretical molecular weigh at 18517 Da and a pI at 6.8. The observed weight in a SDS-PAGE is 22 kDa.

In an 182 amino acids overlap the translated protein has 90,1% identity with E235739; a peptidyl-prolyl cis-trans isomerase.

CFP25

A sequence 93% identical to the 15 determined amino acids was found on the cosmid MTCY339.08c. The one amino acid that differs between the two sequences is a C in MTCY339.08c and a X from the N-terminal sequence data. On a Sequencer a C can not be detected which is a probable explanation for this difference.

The N-terminally determined sequence from the protein purified from culture filtrate begins at amino acid 33 in agreement with the presence of a signal sequence that has been cleaved off. This gives a length of the mature protein of 187 amino acids, which corresponds to a theoretical molecular weigh at 19665 Da, and a theoretical pI at 4.9. The observed weight in a SDS-PAGE is 25 kDa.

In a 217 amino acids overlap the protein has 42.9% identity to CFP21 (MTCY39.35).

CFP28

No homology was found when using the 10 determined amino acid residues 2-8, 11, 12, and 14 of SEQ ID NO: 22 in the database search.

CFP29

Sanger database searching: A sequence nearly 100% identical to the 150 bp sequence of the CFP29 protein was found on cosmid cy444. The sequence is contained within a 795 bp open reading frame of which the 5' end translates into a sequence that is 100% identical to the N-terminally sequenced 19 amino acids of the purified CFP29 protein. The open reading frame encodes a 265 amino acid protein.

The amino acid analysis performed on the purified protein further confirmed the identity of CFP29 with the protein encoded in open reading frame on cosmid 444.

EMBL database searching: The open reading frame encodes a 265 amino acid protein that is 58% identical and 74% similar to the Linocin M18 protein (61% identity on DNA level). This is a 28.6 kDa protein with bacteriocin activity (Valdés-Stauber and Scherer, 1994; Valdés-Stauber and Scherer, 1996). The two proteins have the same length (except for 1 amino acid) and share the same theoretical physicochemical properties. We therefore suggest that CFP29 is a mycobacterial homolog to the *Brevibacterium linens* Linocin M18 protein.

The amino acid sequences of the purified antigens as picked from the Sanger database are shown in the following list. The amino acids determined by N-terminal sequencing are marked with bold.

CFP17 (SEQ ID NO: 6):

```
   1  MTDMNPDIEK DQTSDEVTVE TTSVFRADFL SELDAPAQAG TESAVSGVEG
  51  LPPGSALLVV KRGPNAGSRF LLDQAITSAG RHPDSDIFLD DVTVSRRHAE
 101  FRLENNEFNV VDVGSLNGTY VNREPVDSAV LANGDEVQIG KFRLVFLTGP
 151  KQGEDDGSTG GP
```

CFP20 (SEQ ID NO: 8):

```
   1 MAQITLRGNA INTVGELPAV GSPAPAFTLT GGDLGVISSD QFRGKSVLLN
  51 IFPSVDTPVC ATSVRTFDER AAASGATVLC VSKDLPFAQK RFCGAEGTEN
 101 VMPASAFRDS FGEDYGVTIA DGPMAGLLAR AIVVIGADGN VAYTELVPEI
 151 AQEPNYEAAL AALGA
```

CFP21 (SEQ ID NO: 10):

```
   1 MTPRSLVRIV GVVVATTLAL VSAPAGGRAA HADPCSDIAV
  41 VFARGTHQAS GLGDVGEAFV DSLTSQVGGR SIGVYAVNYP ASDDYRASAS
  91 NGSDDASAHI QRTVASCPNT RIVLGGYSQG ATVIDLSTSA MPPAVADHVA
```

141 AVALPGEPSS GFSSMLWGGG SLPTIGPLYS SKTINLCAPD DPICTGGGNI
191 MAHVSYVQSG MTSQAATFAA NRLDHAG

CFP22 (SEQ ID NO: 12):

1 MADCDSV**TNS PLATATATLE TN**RGDIKIAL PGNHAPKTVA NFVGLAQGTK
51 DYSTQNASGG PSGPFYDGAV FHRVIQGFMI QGGDPTGTGR GGPGYKFADE
101 FHPELQFDKP YLLAMANAGP GTNGSQFFIT VGKTPHLNRR HTIPGEVIDA
151 ESQRVVEAIS KTATDGNDRP TDPVVIESIT IS

CFP25 (SEQ ID NO: 14):

1 MGAAAAMLAA VLLLTPITVP AGYPGAVAPA **TAACPDAEVV FARGRFE**PPG
51 IGTVGNAFVS ALRSKVNKNV GVYAVKYPAD NQIDVGANDM SAHIQSMANS
101 CPNTRLVPGG YSLGAAVTDV VLAVPTQMWG FTNPLPPGSD EHIAAVALFG
151 NGSQWVGPIT NFSPAYNDRT IELCHGDDFV CHPADPNTWE ANWPQHLAGA
201 YVSSGMVNQA ADFVAGKLQ

CFP29 (SEQ ID NO: 16):

1 **MNNLYRDLAP VTEAAWAEIE** LEAARTFKRH IAGRRVVDVS DPGGPVTAAV
51 STGRLIDVKA PTNGVIAHLR ASKPLVRLRV PFTLSRNRID DVERGSKDSD
101 WEPVKEAAKK LAPVEDRTIF EGYSAASIEG IRSASSNPAL TLPEDPREIP
151 DVISQALSEL RLAGVDGPYS VLLSADVYTK VSETSDHGYP IREHLNRLVD
201 GDIIWAPAID GAFVLTTRGG DPDLQLGTDV AIGYASHDTD TVRLYLQETL
251 TFLCYTAEAS VALSH

For all six proteins the molecular weights predicted from the sequences are in agreement with the molecular weights observed on SDS-PAGE.

Cloning of the genes encoding CFP17, CFP20, CFP21, CFP22 and CFP25.

The genes encoding CFP17, CFP20, CFP21, CFP22 and CFP25 were all cloned into the expression vector pMCT6, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the proteins.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1x low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 µl reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles according to the following program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluescript SK II+ - T vector (Stratagene). Plasmid DNA was thereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidine residues which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

CFP17: Primers used for cloning of cfp17:

OPBR-51:	ACAGATCTGTGACGGACATGAACCCG	(SEQ ID NO: 117)
OPBR-52:	TTTTCCATGGTCACGGGCCCCCGGTACT	(SEQ ID NO: 118)

OPBR-51 and OPBR-52 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

CFP20: Primers used for cloning of cfp20:

OPBR-53: ACAGATCTGTGCCCATGGCACAGATA (SEQ ID NO: 119)
OPBR-54: TTTAAGCTTCTAGGCGCCCAGCGCGGC (SEQ ID NO: 120)

OPBR-53 and OPBR-54 create BglII and HinDIII sites, respectively, used for the cloning in pMCT6.

CFP21: Primers used for cloning of cfp21:

OPBR-55: ACAGATCTGCGCATGCGGATCCGTGT (SEQ ID NO: 121)
OPBR-56: TTTTCCATGGTCATCCGGCGTGATCGAG (SEQ ID NO: 122)

OPBR-55 and OPBR-56 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

CFP22: Primers used for cloning of cfp22:

OPBR-57: ACAGATCTGTAATGGCAGACTGTGAT (SEQ ID NO: 123)
OPBR-58: TTTTCCATGGTCAGGAGATGGTGATCGA (SEQ ID NO: 124)

OPBR-57 and OPBR-58 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

CFP25: Primers used for cloning of cfp25:

OPBR-59: ACAGATCTGCCGGCTACCCCGGTGCC (SEQ ID NO: 125)
OPBR-60: TTTTCCATGGCTATTGCAGCTTTCCGGC (SEQ ID NO: 126)

OPBR-59 and OPBR-60 create BglII and NcoI sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP17, CFP20, CFP21, CFP22 and CFP25 proteins.

Expression and metal affinity purification of recombinant proteins was undertaken essentially as described by the manufacturers. For each protein, 1 l LB-media containing 100

μg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmids. Cultures were shaken at 37 °C until they reached a density of OD_{600} = 0.4 - 0.6. IPTG was hereafter added to a final concentration of 1 mM and the cultures were further incubated 4 - 16 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.
After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analyzed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content were determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 3A

Identification of CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B, CFP22A, CFP23A, CFP23B, CFP25A, CFP27, CFP30A, CWP32 and CFP50.

Identification of CFP16 and CFP19B.

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5 % (w/v) and 5 % (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with a 8M urea buffer containing 0.5 % (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a Prep Cell (BioRad) in a matrix of 16% polyacrylamide under an electrical gradient. Fractions containing well separated bands in SDS-PAGE were selected for N-terminal sequencing after transfer to PVDF membrane.

Isolation of CFP8A, CFP8B, CFP19, CFP23A, and CFP23B.

ST-CF was precipitated with ammonium sulphate at 80% saturation and redissolved in PBS, pH 7.4, and dialysed 3 times against 25mM Piperazin-HCl, pH 5.5, and subjected to chromatofocusing on a matrix of PBE 94 (Pharmacia) in a column connected to an FPLC system (Pharmacia). The column was equilibrated with 25 mM Piperazin-HCl, pH 5.5, and the elution was performed with 10% PB74-HCl, pH 4.0 (Pharmacia).

Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml and separated on a Prepcell as described above.

Identification of CFP22A

ST-CF was concentrated approximately 10 fold by ultrafiltration and proteins were precipitated at 80 % saturation, redissolved in PBS, pH 7.4, and dialysed 3 times against PBS, pH 7.4. 5.1 ml of the dialysed ST-CF was treated with RNase (0.2 mg/ml, QUIAGEN) and DNase (0.2 mg/ml, Boehringer Mannheim) for 6 h and placed on top of 6.4 ml of 48 % (w/v) sucrose in PBS, pH 7.4, in Sorvall tubes (Ultracrimp 03987, DuPont Medical Products) and ultracentrifuged for 20 h at 257,300 × g_{max}, 10°C. The pellet was redissolved in 200 µl of 25 mM Tris-192 mM glycine, 0.1 % SDS, pH 8.3.

Identification of CFP7A, CFP25A, CFP27, CFP30A and CFP50

For CFP27, CFP30A and CFP50 ST-CF was concentrated approximately 10 fold by ultrafiltration and ammonium sulphate precipitation in the 45 to 55 % saturation range was performed. Proteins were redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography on an Affi-T gel column (Kem-En-Tec). Proteins were eluted by a 1.5 to 0 M decreasing gradient of ammonium sulphate. Fractions with similar band patterns in SDS-PAGE were pooled and anion exchange chromatography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5, and the elution was performed with a gradient of NaCl from 0 to 1 M. Fractions containing well separated bands in SDS-PAGE were selected.

CFP7A and CFP25A were obtained as described above except for the following modification: ST-CF was concentrated approximately 10 fold by ultrafiltration and proteins were precipi-

tated at 80 % saturation, redissolved in PBS, pH 7.4, and dialysed 3 times against PBS, pH 7.4. Ammonium sulphate was added to a concentration of 1.5 M, and ST-CP proteins were loaded on an Affi T-gel column. Elution from the Affi T-gel column and anion exchange were performed as described above.

Isolation of CWP32

Heat treated H37Rv was subfractionated into subcellular fractions as described in Sørensen et al 1995. The Cell wall fraction was resuspended in 8 M urea, 0.2 % (w/v) N-octyl β-D-glucopyranoside (Sigma) and 5 % (v/v) glycerol and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad) which was equilibrated with the same buffer. Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed by SDS-PAGE and fractions containing well separated bands were polled and subjected to N-terminal sequencing after transfer to PVDF membrane.

N-terminal sequencing

Fractions containing CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B, CFP22A, CFP23A, CFP23B, CFP27, CFP30A, CWP32, and CFP50A were blotted to PVDF membrane after Tricine SDS-PAGE (Ploug et al, 1989). The relevant bands were excised and subjected to N-terminal amino acid sequence analysis on a Procise 494 sequencer (Applied Biosystems). The fraction containing CFP25A was blotted to PVDF membrane after 2-DE PAGE (isoelectric focusing in the first dimension and Tricin SDS-PAGE in the second dimension). The relevant spot was excised and sequenced as described above.

The following N-terminal sequences were obtained:

CFP7A:	AEDVRAEIVA SVLEVVVNEG DQIDKGDVVV LLESMYMEIP VLAEAAGTVS	(SEQ ID NO: 81)
CFP8A:	DPVDDAFIAKLNTAG	(SEQ ID NO: 73)
CFP8B:	DPVDAIINLDNYGX	(SEQ ID NO: 74)

CFP16:	AKLSTDELLDAPKEM	(SEQ ID NO: 79)
CFP19:	TTSPDPYAALPKLPS	(SEQ ID NO: 82)
CFP19B:	DPAXAPDVPTAAQLT	(SEQ ID NO: 80)
CFP22A:	TEYEGPKTKF HALMQ	(SEQ ID NO: 83)
CFP23A:	VIQ/AGMVT/GHIHXVAG	(SEQ ID NO: 76)
CFP23B:	AEMKXFKNAIVQEID	(SEQ ID NO: 75)
CFP25A:	AIEVSVLRVF TDSDG	(SEQ ID NO: 78)
CWP32:	TNIVVLIKQVPDTWS	(SEQ ID NO: 77)
CFP27:	TTIVALKYPG GVVMA	(SEQ ID NO: 84)
CFP30A:	SFPYFISPEX AMRE	(SEQ ID NO: 85)
CFP50:	THYDVVVLGA GPGGY	(SEQ ID NO: 86)

N-terminal homology searching in the Sanger database and identification of the corresponding genes.

The N-terminal amino acid sequence from each of the proteins was used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* database:

http://www.sanger.ac.uk/projects/m-tuberculosis/TB-blast-server.

For CFP23B, CFP23A, and CFP19B no similarities were found in the Sanger database. This could be due to the fact that only approximately 70% of the *M. tuberculosis* genome had been sequenced when the searches were performed. The genes encoding these proteins could be contained in the remaining 30% of the genome for which no sequence data is yet available.

For CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19B , CFP22A, CFP25A, CFP27, CFP30A, CWP32, and CFP50, the following information was obtained:

CFP7A: Of the 50 determined amino acids in CFP7A a 98% identical sequence was found in cosmid csCY07D1 (contig 256): Score = 226 (100.4 bits), Expect = 1.4e-24, P = 1.4e-24 Identities = 49/50 (98%), Positives = 49/50 (98%), Frame = -1

```
Query: 1       AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESMYMEIPVLAEAAGTVS 50
               AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESM MEIPVLAEAAGTVS
Sbjct: 257679 AEDVRAEIVASVLEVVVNEGDQIDKGDVVVLLESMKMEIPVLAEAAGTVS 257530
```

(SEQ ID NOs: 127, 128, and 129)

The identity is found within an open reading frame of 71 amino acids length corresponding to a theoretical MW of CFP7A of 7305.9 Da and a pI of 3.762. The observed molecular weight in an SDS-PAGE gel is 7 kDa.

CFP8A: A sequence 80% identical to the 15 N-terminal amino acids was found on contig TB_1884. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 32. This gives a length of the mature protein of 98 amino acids corresponding to a theoretical MW of 9700 Da and a pI of 3.72 This is in good agreement with the observed MW on SDS-PAGE at approximately 8 kDa. The full length protein has a theoretical MW of 12989 Da and a pI of 4.38.

CFP8B: A sequence 71% identical to the 14 N-terminal amino acids was found on contig TB_653. However, careful re-evaluation of the original N-terminal sequence data confirmed the identification of the protein. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 29. This gives a length of the mature protein of 82 amino acids corresponding to a theoretical MW of 8337 Da and a pI of 4.23. This is in good agreement with the observed MW on SDS-PAGE at approximately 8 kDa. Analysis of the amino acid sequence predicts the presence of a signal peptide which has been cleaved of the mature protein found in culture filtrate.

CFP16: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY20H1.

The identity is found within an open reading frame of 130 amino acids length corresponding to a theoretical MW of CFP16

of 13440.4 Da and a pI of 4.59. The observed molecular weight in an SDS-PAGE gel is 16 kDa.

CFP19: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY270.

The identity is found within an open reading frame of 176 amino acids length corresponding to a theoretical MW of CFP19 of 18633.9 Da and a pI of 5.41. The observed molecular weight in an SDS-PAGE gel is 19 kDa.

CFP22A: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY1A6.

The identity is found within an open reading frame of 181 amino acids length corresponding to a theoretical MW of CFP22A of 20441.9 Da and a pI of 4.73. The observed molecular weight in an SDS-PAGE gel is 22 kDa.

CFP25A: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on contig 255.

The identity is found within an open reading frame of 228 amino acids length corresponding to a theoretical MW of CFP25A of 24574.3 Da and a pI of 4.95. The observed molecular weight in an SDS-PAGE gel is 25 kDa.

CFP27: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found on cosmid MTCY261.

The identity is found within an open reading frame of 291 amino acids length. The N-terminally determined sequence from the protein purified from culture filtrate starts at amino acid 58. This gives a length of the mature protein of 233 amino acids, which corresponds to a theoretical molecular weigh at 24422.4 Da, and a theoretical pI at 4.64. The observed weight in an SDS-PAGE gel is 27 kDa.

CFP30A: Of the 13 determined amino acids in CFP30A, a 100% identical sequence was found on cosmid MTCY261.

The identity is found within an open reading frame of 248 amino acids length corresponding to a theoretical MW of CFP30A of 26881.0 Da and a pI of 5.41. The observed molecular weight in an SDS-PAGE gel is 30 kDa.

CWP32: The 15 amino acid N-terminal sequence was found to be 100% identical to a sequence found on contig 281. The identity was found within an open reading frame of 266 amino acids length, corresponding to a theoretical MW of CWP32 of 28083 Da and a pI of 4.563. The observed molecular weight in an SDS-PAGE gel is 32 kDa.

CFP50: The 15 aa N-terminal sequence was found to be 100% identical to a sequence found in MTV038.06. The identity is found within an open reading frame of 464 amino acids length corresponding to a theoretical MW of CFP50 of 49244 Da and a pI of 5.66. The observed molecular weight in an SDS-PAGE gel is 50 kDa.

Use of homology searching in the EMBL database for identification of CFP19A and CFP21.

Homology searching in the EMBL database (using the GCG package of the Biobase, Århus-DK) with the amino acid sequences of two earlier identified highly immunoreactive ST-CF proteins, using the TFASTA algorithm, revealed that these proteins (CFP21 and CFP25, EXAMPLE 3) belong to a family of fungal cutinase homologs. Among the most homologous sequences were also two *Mycobacterium tuberculosis* sequences found on cosmid MTCY13E12. The first, MTCY13E12.04 has 46% and 50% identity to CFP25 and CFP21 respectively. The second, MTCY13E12.05, has also 46% and 50% identity to CFP25 and CFP21. The two proteins share 62.5% aa identity in a 184 residues overlap. On the basis of the high homology to the strong T-cell antigens CFP21 and CFP25, respectively, it is

believed that CFP19A and CFP23 are possible new T-cell antigens.

The first reading frame encodes a 254 amino acid protein of which the first 26 aa constitute a putative leader peptide that strongly indicates an extracellular location of the protein. The mature protein is thus 228 aa in length corresponding to a theoretical MW of 23149.0 Da and a Pi of 5.80. The protein is named CFP23.

The second reading frame encodes an 231 aa protein of which the first 44 aa constitute a putative leader peptide that strongly indicates an extracellular location of the protein. The mature protein is thus 187 aa in length corresponding to a theoretical MW of 19020.3 Da and a Pi of 7.03. The protein is named CFP19A.

The presence of putative leader peptides in both proteins (and thereby their presence in the ST-CF) is confirmed by theoretical sequence analysis using the signalP program at the Expasy molecular Biology server

(http://expasy.hcuge.ch/www/tools.html).

Searching for homologies to CFP7A, CFP16, CFP19, CFP19A, CFP19B, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32 and CFP50 in the EMBL database.

The amino acid sequences derived from the translated genes of the individual antigens were used for homology searching in the EMBL and Genbank databases using the TFASTA algorithm, in order to find homologous proteins and to address eventual functional roles of the antigens.

CFP7A: CFP7A has 44% identity and 70% similarity to hypothetical *Methanococcus jannaschii* protein (*M. jannaschii* from base 1162199-1175341), as well as 43% and 38% identity and 68 and 64% similarity to the C-terminal part of *B. stearothermo-*

philus pyruvate carboxylase *and Streptococcus mutans* biotin carboxyl carrier protein.

CFP7A contains a consensus sequence EAMKM for a biotin binding site motif which in this case was slightly modified (ESMKM in amino acid residues 34 to 38). By incubation with alkaline phosphatase conjugated streptavidin after SDS-PAGE and transfer to nitrocellulose it was demonstrated that native CFP7A was biotinylated.

CFP16: RplL gene, 130 aa. Identical to the *M. bovis* 50s ribosomal protein L7/L12 (acc. No P37381).

CFP19: CFP19 has 47% identity and 55% similarity to *E.coli* pectinesterase homolog (ybhC gene) in a 150 aa overlap.

CFP19A: CFP19A has between 38% and 45% identity to several cutinases from different fungal sp.

In addition CFP19A has 46% identity and 61% similarity to CFP25 as well as 50% identity and 64% similarity to CFP21 (both proteins are earlier isolated from the ST-CF).

CFP19B: No apparent homology

CFP22A: No apparent homology

CFP23: CFP23 has between 38% and 46% identity to several cutinases from different fungal sp.

In addition CFP23 has 46% identity and 61% similarity to CFP25 as well as 50% identity and 63% similarity to CFP21 (both proteins are earlier isolated from the ST-CF).

CFP25A: CFP25A has 95% identity in a 241 aa overlap to a putative *M. tuberculosis* thymidylate synthase (450 aa accession No p28176).

CFP27: CFP27 has 81% identity to a hypothetical *M. leprae* protein and 64% identity and 78% similarity to *Rhodococcus* sp. proteasome beta-type subunit 2 (prcB(2) gene).

CFP30A: CFP30A has 67% identity to *Rhodococcus proteasome* alfa-type 1 subunit.

CWP32: The CWP32 N-terminal sequence is 100% identical to the *Mycobacterium leprae* sequence MLCB637.03.

CFP50: The CFP50 N-terminal sequence is 100% identical to a putative lipoamide dehydrogenase from *M. leprae* (Accession 415183)

Cloning of the genes encoding CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32, and CFP50.

The genes encoding CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32 and CFP50 were all cloned into the expression vector pMCT6, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the proteins.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1X low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 ml reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec. using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluescript SK II+ - T vector (Stratagene). Plasmid DNA was hereafter prepared from clones

harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidines which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

CFP7A: Primers used for cloning of *cfp7A*:

OPBR-79:	AAGAGTAGATCTATGATGGCGAGGATGTTCGCG	(SEQ ID NO: 95)
OPBR-80:	CGGCGACGACGGATCCTACCGCGTCGG	(SEQ ID NO: 96)

OPBR-79 and OPBR-80 create *Bgl*II and *Bam*HI sites, respectively, used for the cloning in pMCT6.

CFP8A: Primers used for cloning of *cfp8A*:

CFP8A-F:	CTGAGATCTATGAACCTACGGCGCC	(SEQ ID NO: 154)
CFP8A-R:	CTCCCATGGTACCCTAGGACCCCGGCAGCCCCGGC	(SEQ ID NO: 155)

CFP8A-F and CFP8A-R create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP8B: Primers used for cloning of *cfp8B*:

CFP8B-F:	CTGAGATCTATGAGGCTGTCGTTGACCGC	(SEQ ID NO: 156)
CFP8B-R:	CTCCCCGGGCTTAATAGTTGTTGCAGGAGC	(SEQ ID NO: 157)

CFP8B-F and CFP8B-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP16: Primers used for cloning of *cfp*16:

OPBR-104:	CCGGGAGATCTATGGCAAAGCTCTCCACCGACG	(SEQ ID NOs: 111 and 130)
OPBR-105:	CGCTGGGCAGAGCTACTTGACGGTGACGGTGG	(SEQ ID NOs: 112 and 131)

OPBR-104 and OPBR-105 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP19: Primers used for cloning of *cfp*19:

OPBR-96:	GAGGAAGATCTATGACAACTTCACCCGACCCG	(SEQ ID NO: 107)
OPBR-97.	CATGAAGCCATGGCCCGCAGGCTGCATG	(SEQ ID NO: 108)

OPBR-96 and OPBR-97 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP19A: Primers used for cloning of *cfp*19A:

OPBR-88:	CCCCCCAGATCTGCACCACCGGCATCGGCGGGC	(SEQ ID NO: 99)
OPBR-89.	GCGGCGGATCCGTTGCTTAGCCGG	(SEQ ID NO: 100)

OPBR-88 and OPBR-89 create *Bgl*II and *Bam*HI sites, respectively, used for the cloning in pMCT6.

CFP22A: Primers used for cloning of *cfp*22A:

OPBR-90:	CCGGCTGAGATCTATGACAGAATACGAAGGGC	(SEQ ID NO: 101)
OPBR-91:	CCCCGCCAGGGAACTAGAGGCGGC	(SEQ ID NO: 102)

OPBR-90 and OPBR-91 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP23: Primers used for cloning of *cfp*23:

OPBR-86:	CCTTGGGAGATCTTTGGACCCCGGTTGC	(SEQ ID NO: 97)
OPBR-87:	GACGAGATCTTATGGGCTTACTGAC	(SEQ ID NO: 98)

OPBR-86 and OPBR-87 both create a *Bgl*II site used for the cloning in pMCT6.

CFP25A: Primers used for cloning of *cfp*25A:

OPBR-106:	GGCCCAGATCTATGGCCATTGAGGTTTCGGTGTTGC	(SEQ ID NO: 113)
OPBR-107:	CGCCGTGTTGCATGGCAGCGCTGAGC	(SEQ ID NO: 114)

OPBR-106 and OPBR-107 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP27: Primers used for cloning of *cfp*27:

OPBR-92:	CTGCCGAGATCTACCACCATTGTCGCGCTGAAATACCC	(SEQ ID NO: 103)
OPBR-93:	CGCCATGGCCTTACGGCGCAACTCG	(SEQ ID NO: 104)

OPBR-92 and OPBR-93 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP30A: Primers used for cloning of *cfp*30A:

OPBR-94:	GGCGGAGATCTGTGAGTTTTCCGTATTTCATC	(SEQ ID NO: 105)
OPBR-95:	CGCGTCGAGCCATGGTTAGGCGCAG	(SEQ ID NO: 106)

OPBR-94 and OPBR-95 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CWP32: Primers used for cloning of *cwp*32:

CWP32-F:	GCTTAGATCTATGATTTTCTGGGCAACCAGGTA	(SEQ ID NO: 158)
CWP32-R:	GCTTCCATGGCGAGGCACAGGCGTGGGAA	(SEQ ID NO: 159)

CWP32-F and CWP32-R create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

CFP50: Primers used for cloning of *cfp*50:

OPBR-100:	GGCCGAGATCTGTGACCCACTATGACGTCGTCG	(SEQ ID NO: 109)
OPBR-101:	GGCGCCCATGGTCAGAAATTGATCATGTGGCCAA	(SEQ ID NO: 110)

OPBR-100 and OPBR-101 create *Bgl*II and *Nco*I sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP7A, CFP8A, CFP8B, CFP16, CFP19, CFP19A, CFP22A, CFP23, CFP25A, CFP27, CFP30A, CWP32, and CFP50 proteins.

Expression and metal affinity purification of recombinant proteins was undertaken essentially as described by the manufacturers. For each protein, 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmids. Cultures were shaken at 37°C until they reached a density of OD_{600} = 0.4 - 0.6. IPTG was hereafter added to a final concentration of 1 mM and the cultures were further incubated 4-16 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analyzed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content were determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 3B

Identification of CFP7B, CFP10A, CFP11 and CFP30B.

Isolation of CFP7B

ST-CF was precipitated with ammonium sulphate at 80% saturation and redissolved in PBS, pH 7.4, and dialyzed 3 times against 25 mM Piperazin-HCl, pH 5.5, and subjected to cromatofocusing on a matrix of PBE 94 (Pharmacia) in a column connected to an FPLC system (Pharmacia). The column was equilibrated with 25 mM Piperazin-HCl, pH 5.5, and the elution was performed with 10% PB74-HCl, pH 4.0 (Pharmacia). Fractions with similar band patterns were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off membrane to a final volume of 1-3 ml. An equal volume of SDS containing sample buffer was added and the protein solution boiled for 5 min before further separation on a MultiEluter (BioRad) in a matrix of 10-20 % polyacrylamid (Andersen,P. & Heron,I., 1993). The fraction containing a well separated band below 10 kDa was selected for N-terminal sequencing after transfer to a PVDF membrane.

Isolation of CFP11

ST-CF was precipitated with ammonium sulphate at 80% saturation. The precipitated proteins were removed by centrifugation and after resuspension washed with 8 M urea. CHAPS and glycerol were added to a final concentration of 0.5 % (w/v) and 5% (v/v) respectively and the protein solution was applied to a Rotofor isoelectrical Cell (BioRad). The Rotofor Cell had been equilibrated with an 8M urea buffer containing 0.5 % (w/v) CHAPS, 5% (v/v) glycerol, 3% (v/v) Biolyt 3/5 and 1% (v/v) Biolyt 4/6 (BioRad). Isoelectric focusing was performed in a pH gradient from 3-6. The fractions were analyzed on silver-stained 10-20% SDS-PAGE. The fractions in the pH gradient 5.5 to 6 were pooled and washed three times with PBS on a Centriprep concentrator (Amicon) with a 3 kDa cut off

membrane to a final volume of 1 ml. 300 mg of the protein preparation was separated on a 10-20% Tricine SDS-PAGE (Ploug et al 1989) and transferred to a PVDF membrane and Coomassie stained. The lowest band occurring on the membrane was excised and submitted for N-terminal sequencing.

Isolation of CFP10A and CFP30B

ST-CF was concentrated approximately 10-fold by ultrafiltration and ammonium sulphate precipitation at 80 % saturation. Proteins were redissolved in 50 mM sodium phosphate, 1.5 M ammonium sulphate, pH 8.5, and subjected to thiophilic adsorption chromatography on an Affi-T gel column (Kem-En-Tec). Proteins were eluted by a 1.5 to 0 M decreasing gradient of ammonium sulphate. Fractions with similar band patterns in SDS-PAGE were pooled and anion exchange chromatography was performed on a Mono Q HR 5/5 column connected to an FPLC system (Pharmacia). The column was equilibrated with 10 mM Tris-HCl, pH 8.5, and the elution was performed with a gradient of NaCl from 0 to 1 M. Fractions containing well separated bands in SDS-PAGE were selected.

Fractions containing CFP10A and CFP30B were blotted to PVDF membrane after 2-DE PAGE (Ploug et al, 1989). The relevant spots were excised and subjected to N-terminal amino acid sequence analysis.

N-terminal sequencing

N-terminal amino acid sequence analysis was performed on a Procise 494 sequencer (applied Biosystems).

The following N-terminal sequences were obtained:

CFP7B:	PQGTVKWFNAEKGFG	(SEQ ID NO: 168)
CFP10A:	NVTVSIPTILRPXXX	(SEQ ID NO: 169)
CFP11:	TRFMTDPHAMRDMAG	(SEQ ID NO: 170)
CFP30B:	PKRSEYRQGTPNWVD	(SEQ ID NO: 171)

"X" denotes an amino acid which could not be determined by the sequencing method used.

N-terminal homology searching in the Sanger database and identification of the corresponding genes.

The N-terminal amino acid sequence from each of the proteins was used for a homology search using the blast program of the Sanger *Mycobacterium tuberculosis* genome database:

http://www.sanger.ac.uk/projects/m-tuberculosis/TB-blast-server.

For CFP11 a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_1314. The identity was found within an open reading frame of 98 amino acids length corresponding to a theoretical MW of 10977 Da and a pI of 5.14.

Amino acid number one can also be an Ala (insted of a Thr) as this sequence was also obtained (results not shown), and a 100% identical sequence to this N-terminal is found on contig TB_671 and on locus MTCI364.09.

For CFP7B a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_2044 and on locus MTY15C10.04 with EMBL accession number: z95436. The identity was found within an open reading frame of 67 amino acids length corresponding to a theoretical MW of 7240 Da and a pI of 5.18.

For CFP10A a sequence 100% identical to 12 N-terminal amino acids was found on contig TB_752 and on locus CY130.20 with EMBL accession number: Q10646 and Z73902. The identity was found within an open reading frame of 93 amino acids length corresponding to a theoretical MW of 9557 Da and a pI of 4.78.

For CFP30B a sequence 100% identical to 15 N-terminal amino acids was found on contig TB_335. The identity was found within an open reading frame of 261 amino acids length

corresponding to a theoretical MW of 27345 Da and a pI of 4.24.

The amino acid sequences of the purified antigens as picked from the Sanger database are shown in the following list.

CFP7B (SEQ ID NO: 147)

```
  1  MPQGTVKWFN AEKGFGFIAP EDGSADVFVH YTEIQGTGFR TLEENQKVEF
 51  EIGHSPKGPQ ATGVRSL
```

CFP10A (SEQ ID NO: 141)

```
  1  MNVTVSIPTI LRPHTGGQKS VSASGDTLGA VISDLEANYS GISERLMDPS
 51  SPGKLHRFVN IYVNDEDVRF SGGLATAIAD GDSVTILPAV AGG
```

CFP11 protein sequence (SEQ ID NO: 143)

```
  1  MATRFMTDPH AMRDMAGRFE VHAQTVEDSA RRMWASAQNI SGAGWSGMAS
 51  ATSLDTMAQM NQAFRNIVNM LHGVRDGLVR DANNYEQQEQ ASQQILSS
```

CFP30B (SEQ ID NO: 145)

```
  1  MPKRSEYRQG TPNWVDLQTT DQSAAKKFYT SLFGWGYDDN PVPGGGGVYS
 51  MATLNGEAVA AIAPMPPGAP EGMPPIWNTY IAVDDVDAVV DKVVPGGGQV
101  MMPAPDIGDA GRMSFITDPT GAAVGLWQAN RHIGATLVNE TGTLIWNELL
151  TDKPDLALAF YEAVVGLTHS SMEIAAGQNY RVLKAGDAEV GGCMEPPMPG
201  VPNHWNVYFA VDDADATAAK AAAAGGQVIA EPADIPSVGR FAVLSDPQGA
251  IFSVLKPAPQ Q
```

Cloning of the genes encoding CFP7B, CFP10A, CFP11, and CFP30B.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1X low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stra-

tagene) in 10 ml reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec., using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns (Costar) and cloned into pBluscript SK II+ - T vector (Stratagene). Plasmid DNA was hereafter prepared from clones harbouring the desired fragments, digested with suitable restriction enzymes and subcloned into the expression vector pMCT6 in frame with 8 histidines which are added to the N-terminal of the expressed proteins. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

For cloning of the individual antigens, the following gene specific primers were used:

CFP7B: Primers used for cloning of *cfp*7B:

CFP7B-F:	CTGAGATCTAGAATGCCACAGGGAACTGTG	(SEQ ID NO: 160)
CFP7B-R:	TCTCCCGGGGGTAACTCAGAGCGAGCGGAC	(SEQ ID NO: 161)

CFP7B-F and CFP7B-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP10A: Primers used for cloning of *cfp*10A:

CFP10A-F:	CTGAGATCTATGAACGTCACCGTATCC	(SEQ ID NO: 162)
CFP10A-R:	TCTCCCGGGGCTCACCCACCGGCCACG	(SEQ ID NO: 163)

CFP10A -F and CFP10A -R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP11: Primers used for cloning of *cfp*11:

CFP11-F:	CTGAGATCTATGGCAACACGTTTTATGACG	(SEQ ID NO: 164)
CFP11-R:	CTCCCCGGGTTAGCTGCTGAGGATCTGCTH	(SEQ ID NO: 165)

CFP11-F and CFP11-R create *Bgl*II and *Sma*I sites, respectively, used for the cloning in pMCT6.

CFP30B: Primers used for cloning of *cfp*30B:

CFP30B-F:	CTGAAGATCTATGCCCAAGAAAAGCGAATAC	(SEQ ID NO: 166)
CFP30B -R:	CGGCAGCTGCTAGCATTCTCCGAATCTGCCG	(SEQ ID NO: 167)

CFP30B-F and CFP30B-R create *Bgl*II and *Pvu*II sites, respectively, used for the cloning in pMCT6.

Expression/purification of recombinant CFP7B, CFP10A, CFP11 and CFP30B protein.

Expression and metal affinity purification of recombinant protein was undertaken essentially as described by the manufacturers. 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmid. The culture was shaken at 37 °C until it reached a density of OD_{630} = 0.5. IPTG was hereafter added to a final concentration of 1 mM and the culture was further incubated 4 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations

were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analysed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content was determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

EXAMPLE 4

Cloning of the gene expressing CFP26 (MPT51)

Synthesis and design of probes

Oligonucleotide primers were synthesized automatically on a DNA synthesizer (Applied Biosystems, Forster City, Ca, ABI-391, PCR-mode) deblocked and purified by ethanol precipitation.

Three oligonucleotides were synthesized (TABLE 3) on the basis of the nucleotide sequence from *mpb51* described by Ohara *et al.* (1995). The oligonucleotides were engineered to include an *Eco*RI restriction enzyme site at the 5' end and at the 3' end by which a later subcloning was possible.

Additional four oligonucleotides were synthesized on the basis of the nucleotide sequence from MPT51 (Fig. 5 and SEQ ID NO: 41). The four combinations of the primers were used for the PCR studies.

DNA cloning and PCR technology

Standard procedures were used for the preparation and handling of DNA (Sambrook *et al.*, 1989). The gene *mpt51* was cloned from *M. tuberculosis* H37Rv chromosomal DNA by the use of the polymerase chain reactions (PCR) technology as described previously (Oettinger and Andersen, 1994). The PCR product was cloned in the pBluescriptSK + (Stratagene).

Cloning of *mpt51*

The gene, the signal sequence and the Shine Delgarno region of MPT51 was cloned by use of the PCR technology as two fragments of 952 bp and 815 bp in pBluescript SK +, designated pTO52 and pTO53.

DNA Sequencing

The nucleotide sequence of the cloned 952 bp *M. tuberculosis* H37Rv PCR fragment, pTO52, containing the Shine Dalgarno sequence, the signal peptide sequence and the structural gene of MPT51, and the nucleotide sequence of the cloned 815 bp PCR fragment containing the structural gene of MPT51, pTO53, were determined by the dideoxy chain termination method adapted for supercoiled DNA by use of the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., Cleveland, OH) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A; Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

The nucleotide sequences of pTO52 and pTO53 and the deduced amino acid sequence are shown in Figure 5. The DNA sequence contained an open reading frame starting with a ATG codon at position 45 - 47 and ending with a termination codon (TAA) at position 942 - 944. The nucleotide sequence of the first 33 codons was expected to encode the signal sequence. On the basis of the known N-terminal amino acid sequence (Ala - Pro

- Tyr - Glu - Asn) of the purified MPT51 (Nagai *et al.*, 1991) and the features of the signal peptide, it is presumed that the signal peptidase recognition sequence (Ala-X-Ala) (von Heijne, 1984) is located in front of the N-terminal region of the mature protein at position 144. Therefore, a structural gene encoding MPT51, *mpt51*, derived from *M. tuberculosis* H37Rv was found to be located at position 144 - 945 of the sequence shown in Fig. 5. The nucleotide sequence of *mpt51* differed with one nucleotide compared to the nucleotide sequence of MPB51 described by Ohara *et al.* (1995) (Fig. 5). In *mpt51* at position 780 was found a substitution of a guanine to an adenine. From the deduced amino acid sequence this change occurs at a first position of the codon giving a amino acid change from alanine to threonine. Thus it is concluded, that *mpt51* consists of 801 bp and that the deduced amino acid sequence contains 266 residues with a molecular weight of 27,842, and MPT51 show 99,8% identity to MPB51.

Subcloning of *mpt51*

An *Eco*RI site was engineered immediately 5' of the first codon of *mpt51* so that only the coding region of the gene encoding MPT51 would be expressed, and an *Eco*RI site was incorporated right after the stop codon at the 3' end.

DNA of the recombinant plasmid pTO53 was cleaved at the *Eco*RI sites. The 815 bp fragment was purified from an agarose gel and subcloned into the *Eco*RI site of the pMAL-*cRI* expression vector (New England Biolabs), pTO54. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue by the standard procedures for DNA manipulation.

The endpoints of the gene fusion were determined by the dideoxy chain termination method as described under section DNA sequencing. Both strands of the DNA were sequenced.

Preparation and purification of rMPT51

Recombinant antigen was prepared in accordance with instructions provided by New England Biolabs. Briefly, single colonies of *E. coli* harbouring the pTO54 plasmid were inoculated into Luria-Bertani broth containing 50 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to 2 x 10^8 cells/ml. Isopropyl-β-D-thiogalactoside (IPTG) was then added to a final concentration of 0.3 mM and growth was continued for further 2 hours. The pelleted bacteria were stored overnight at -20°C in new column buffer (20 mM Tris/HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT))and thawed at 4°C followed by incubation with 1 mg/ml lysozyme on ice for 30 min and sonication (20 times for 10 sec with intervals of 20 sec). After centrifugation at 9,000 x g for 30 min at 4°C, the maltose binding protein-MPT51fusion protein (MBP-rMPT51) was purified from the crude extract by affinity chromatography on amylose resin column. MBP-rMPT51 binds to amylose. After extensive washes of the column, the fusion protein was eluted with 10 mM maltose. Aliquots of the fractions were analyzed on 10% SDS-PAGE. Fractions containing the fusion protein of interest were pooled and was dialysed extensively against physiological saline.

Protein concentration was determined by the BCA method supplied by Pierce (Pierce Chemical Company, Rockford, IL).

TABLE 3.

Sequence of the *mpt51* oligonucleotides[a].

Orientation and oligonucleotide[a]	Sequences (5'→ 3')	Position[b] (nucleotide)
Sense		
MPT51-1	CTCGAATTCGCCGGGTGCACACAG (SEQ ID NO: 28)	6 - 21 (SEQ ID NO: 41)
MPT51-3	CTCGAATTCGCCCCATACGAGAAC (SEQ ID NO: 29)	143 - 156 (SEQ ID NO: 41)
MPT51-5	GTGTATCTGCTGGAC (SEQ ID NO: 30)	228 - 242 (SEQ ID NO: 41)
MPT51-7	CCGACTGGCTGGCCG (SEQ ID NO: 31)	418 - 432 (SEQ ID NO: 41)
Antisense		
MPT51-2	GAGGAATTCGCTTAGCGGATCGCA (SEQ ID NO: 32)	946 - 932 (SEQ ID NO: 41)
MPT51-4	CCCACATTCCGTTGG (SEQ ID NO: 33)	642 - 628 (SEQ ID NO: 41)
MPT51-6	GTCCAGCAGATACAC (SEQ ID NO: 34)	242 - 228 (SEQ ID NO: 41)

[a] The oligonucleotides MPT51-1 and MPT51-2 were constructed from the MPB51 nucleotide sequence (Ohara *et al.*, 1995). The other oligonucleotides constructions were based on the nucleotide sequence obtained from *mpt51* reported in this work. Nucleotides (nt) underlined are not contained in the nucleotide sequence of MPB/T51.

[b] The positions referred to are of the non-underlined parts of the primers and correspond to the nucleotide sequence shown in SEQ ID NO: 41.

Cloning of *mpt51* in the expression vector pMST24.

A PCR fragment was produced from pTO52 using the primer combination MPT51-F and MPT51-R (TABLE 4). A *Bam*HI site was engineered immediately 5' of the first codon of *mpt51* so that only the coding region of the gene encoding MPT51 would be expressed, and an *Nco*I site was incorporated right after the stop codon at the 3' end.

The PCR product was cleaved at the *Bam*HI and the *Nco*I site. The 811 bp fragment was purified from an agarose gel and subcloned into the *Bam*HI and the *Nco*I site of the pMST24 expression vector, pTO86. Vector DNA containing the gene fusion was used to transform the *E. coli* XL1-Blue by the standard procedures for DNA manipulation.

The nucleotide sequence of complete gene fusion was determined by the dideoxy chain termination method as described under section DNA sequencing. Both strands of the DNA were sequenced.

Preparation and purification of rMPT51.

Recombinant antigen was prepared from single colonies of *E. coli* harbouring the pTO86 plasmid inoculated into Luria-Bertani broth containing 50 µg/ml ampicillin and 12.5 µg/ml tetracycline and grown at 37°C to 2 x 10^8 cells/ml. Isopropyl-ß-D-thiogalactoside (IPTG) was then added to a final concentration of 1 mM and growth was continued for further 2 hours. The pelleted bacteria were resuspended in BC 100/20 buffer (100 mM KCl, 20 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol). Cells were broken by sonication (20 times for 10 sec with intervals of 20 sec). After centrifugation at 9,000 x g for 30 min. at 4°C the insoluble matter was resuspended in BC 100/20 buffer with 8 M urea followed by sonication and centrifugation as above. The 6 x His tag-MPT51 fusion protein (His-rMPT51) was purified by affinity chromatography on Ni-NTA resin column (Qiagen, Hilden, Germany). His-rMPT51 binds to Ni-NTA. After extensive washes of the column, the fusion protein was eluted with BC 100/40 buffer (100 mM KCl, 40 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol) with 8 M urea and BC 1000/40 buffer (1000 mM KCl, 40 mM Imidazole, 20 mM Tris/HCl, pH 7.9, 20 % glycerol) with 8 M urea. His-rMPT51 was extensive dialysed against 10 mM Tris/HCl, pH 8.5, 3 M urea followed by purification using fast protein liquid chromatography (FPLC) (Pharmacia, Uppsala, Sweden), over an anion exchange column (Mono Q) using 10 mM Tris/HCl, pH 8.5, 3 M urea with a 0 - 1 M NaCl linear gradient. Fractions containing rMPT51 were pooled and subsequently dialysed extensively against 25 mM Hepes, pH 8.0 before use.

Protein concentration was determined by the BCA method supplied by Pierce (Pierce Chemical Company, Rockford, IL). The lipopolysaccharide (LPS) content was determined by the limulus amoebocyte lysate test (LAL) to be less than 0.004 ng/µg rMPT51, and this concentration had no influence on cellular activity.

TABLE 4. Sequence of the *mpt51* oligonucleotides.

Orientation and oligonucleotide	Sequences (5' → 3')	Position (nt)
Sense		
MPT51-F	CTCGGATCCTGCCCCATACGAGAACCTG	139 - 156
Antisense		
MPT51-R	CTCCCATGGTTAGCGGATCGCACCG	939 - 924

EXAMPLE 4A

Cloning of the ESAT6-MPT59 and the MPT59-ESAT6 hybrides.

Background for ESAT-MPT59 and MPT59-ESAT6 fusion

Several studies have demonstrated that ESAT-6 is a an immunogen which is relatively difficult to adjuvate in order to obtain consistent results when immunizing therewith. To detect an *in vitro* recognition of ESAT-6 after immunization with the antigen is very difficult compared to the strong recognition of the antigen that has been found during the recall of memory immunity to *M. tuberculosis*. ESAT-6 has been found in ST-CF in a truncated version were amino acids 1-15 have been deleted. The deletion includes the main T-cell epitopes recognized by C57BL/6j mice (Brandt et al., 1996). This result indicates that ESAT-6 either is N-terminally processed or proteolytically degraded in STCF. In order to optimize ESAT-6 as an immunogen, a gene fusion between ESAT-6 and another major T cell antigen MPT59 has been constructed. Two different construct have been made: MPT59-ESAT-6 (SEQ ID NO: 172) and ESAT-6-MPT59 (SEQ ID NO: 173). In the first hybrid ESAT-6 is N-terminally protected by MPT59 and in the latter it is expected that the fusion of two dominant T-cell antigens can have a synergistic effect.

The genes encoding the ESAT6-MPT59 and the MPT59-ESAT6 hybrides were cloned into the expression vector pMCT6, by PCR amplification with gene specific primers, for recombinant expression in *E. coli* of the hybrid proteins.

Construction of the hybrid MPT59-ESAT6.

The cloning was carried out in three steps. First the genes encoding the two components of the hybrid, ESAT6 and MPT59, were PCR amplified using the following primer constructions:

ESAT6:

OPBR-4: GGCGCCGGCAAGCTTGCCATGACAGAGCAGCAGTGG (SEQ ID NO: 132)
OPBR-28: CGAACTGGCCGGATCCCCGTGTTTCGC (SEQ ID NO: 133)

OPBR-4 and OPBR-28 create HinDIII and BamHI sites, respectively.

MPT59:

OPBR-48: GGCAACGGCGAGATCTTTCTCCCGGCCGGGC (SEQ ID NO: 134)
OPBR-3: GGCAAGCTTGCCGGCGCCTAACGAACT (SEQ ID NO: 135)

OPBR-48 and OPBR-3 create BglII and HinDIII, respectively. Additionally OPBR-3 deletes the stop codon of MPT59.

PCR reactions contained 10 ng of *M. tuberculosis* chromosomal DNA in 1x low salt Taq+ buffer from Stratagene supplemented with 250 mM of each of the four nucleotides (Boehringer Mannheim), 0,5 mg/ml BSA (IgG technology), 1% DMSO (Merck), 5 pmoles of each primer and 0.5 unit Taq+ DNA polymerase (Stratagene) in 10 µl reaction volume. Reactions were initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 10 sec., 55°C for 10 sec. and 72°C for 90 sec, using thermocycler equipment from Idaho Technology.

The DNA fragments were subsequently run on 1% agarose gels, the bands were excised and purified by Spin-X spin columns

(Costar). The two PCR fragments were digested with HinDIII and ligated. A PCR amplification of the ligated PCR fragments encoding MPT59-ESAT6 was carried out using the primers OPBR-48 and OPBR-28. PCR reaction was initially heated to 94°C for 25 sec. and run for 30 cycles of the program; 94°C for 30 sec., 55°C for 30 sec. and 72°C for 90 sec. The resulting PCR fragment was digested with BglII and BamHI and cloned into the expression vector pMCT6 in frame with 8 histidines which are added to the N-terminal of the expressed protein hybrid. The resulting clones were hereafter sequenced by use of the dideoxy chain termination method adapted for supercoiled DNA using the Sequenase DNA sequencing kit version 1.0 (United States Biochemical Corp., USA) and by cycle sequencing using the Dye Terminator system in combination with an automated gel reader (model 373A, Applied Biosystems) according to the instructions provided. Both strands of the DNA were sequenced.

Construction of the hybrid ESAT6-MPT59.

Construction of the hybrid ESAT6-MPT59 was carried out as described for the hybrid MPT59-ESAT6. The primers used for the construction and cloning were:

ESAT6:

OPBR-75: GGACCCAGATCTATGACAGAGCAGCAGTGG (SEQ ID NO: 136)
OPBR-76: CCGGCAGCCCCGGCCGGGAGAAAAGCTTTGCGAACATCCCAGTGACG (SEQ ID NO: 137)

OPBR-75 and OPBR-76 create BglII and HinDIII sites, respectively. Additionally OPBR-76 deletes the stop codon of ESAT6.

MPT59:

OPBR-77: GTTCGCAAAGCTTTTCTCCCGGCCGGGGCTGCCGGTCGAGTACC (SEQ ID NO: 138)
OPBR-18: CCTTCGGTGGATCCCGTCAG (SEQ ID NO: 139)

OPBR-77 and OPBR-18 create HinDIII and BamHI sites, respectively.

Expression/purification of MPT59-ESAT6 and ESAT6-MPT59 hybrid proteins.

Expression and metal affinity purification of recombinant proteins was undertaken essentially as described by the manufacturers. For each protein, 1 l LB-media containing 100 µg/ml ampicillin, was inoculated with 10 ml of an overnight culture of XL1-Blue cells harbouring recombinant pMCT6 plasmids. Cultures were shaken at 37 °C until they reached a density of OD_{600} = 0.4 - 0.6. IPTG was hereafter added to a final concentration of 1 mM and the cultures were further incubated 4 - 16 hours. Cells were harvested, resuspended in 1X sonication buffer + 8 M urea and sonicated 5 X 30 sec. with 30 sec. pausing between the pulses.

After centrifugation, the lysate was applied to a column containing 25 ml of resuspended Talon resin (Clontech, Palo Alto, USA). The column was washed and eluted as described by the manufacturers.

After elution, all fractions (1.5 ml each) were subjected to analysis by SDS-PAGE using the Mighty Small (Hoefer Scientific Instruments, USA) system and the protein concentrations were estimated at 280 nm. Fractions containing recombinant protein were pooled and dialysed against 3 M urea in 10 mM Tris-HCl, pH 8.5. The dialysed protein was further purified by FPLC (Pharmacia, Sweden) using a 6 ml Resource-Q column, eluted with a linear 0-1 M gradient of NaCl. Fractions were analyzed by SDS-PAGE and protein concentrations were estimated at OD_{280}. Fractions containing protein were pooled and dialysed against 25 mM Hepes buffer, pH 8.5.

Finally the protein concentration and the LPS content were determined by the BCA (Pierce, Holland) and LAL (Endosafe, Charleston, USA) tests, respectively.

The biological activity of the MPT59-ESAT6 fusion protein is described in Example 6A.

EXAMPLE 5

Mapping of the purified antigens in a 2DE system.

In order to characterize the purified antigens they were mapped in a 2-dimensional electrophoresis (2DE) reference system. This consists of a silver stained gel containing ST-CF proteins separated by isoelectrical focusing followed by a separation according to size in a polyacrylamide gel electrophoresis. The 2DE was performed according to Hochstrasser *et al.* (1988). 85 µg of ST-CF was applied to the isoelectrical focusing tubes where BioRad ampholytes BioLyt 4-6 (2 parts) and BioLyt 5-7 (3 parts) were included. The first dimension was performed in acrylamide/piperazin diacrylamide tube gels in the presence of urea, the detergent CHAPS and the reducing agent DTT at 400 V for 18 hours and 800 V for 2 hours. The second dimension 10-20% SDS-PAGE was performed at 100 V for 18 hours and silver stained. The identification of CFP7, CFP7A, CFP7B, CFP8A, CFP8B, CFP9, CFP11, CFP16, CFP17, CFP19, CFP20, CFP21, CFP22, CFP25, CFP27, CFP28, CFP29, CFP30A, CFP50, and MPT51 in the 2DE reference gel were done by comparing the spot pattern of the purified antigen with ST-CF with and without the purified antigen. By the assistance of an analytical 2DE software system (Phoretix International, UK) the spots have been identified in Fig. 6. The position of MPT51 and CFP29 were confirmed by a Western blot of the 2DE gel using the Mab's anti-CFP29 and HBT 4.

EXAMPLE 6

Biological activity of the purified antigens.

<u>IFN-γ induction in the mouse model of TB infection</u>

The recognition of the purified antigens in the mouse model of memory immunity to TB (described in example 1) was investigated. The results shown in TABLE 5 are representative for three experiments.

A very high IFN-γ response was induced by two of the antigens CFP17 and CFP21 at almost the same high level as ST-CF.

TABLE 5

IFN-γ release from splenic memory effector cells from C57BL/6J mice isolated after reinfection with *M. tuberculosis* after stimulation with native antigens.

Antigen[a]	IFN-γ (pg/ml)[b]
ST-CF	12564
CFP7	ND[d]
CFP9	ND
CFP17	9251
CFP20	2388
CFP21	10732
CFP22 + CFP25[c]	5342
CFP26 (MPT51)	ND
CFP28	2818
CFP29	3700

The data is derived from a representative experiment out of three.
[a] ST-CF was tested in a concentration of 5 µg/ml and the individual antigens in a concentration of 2 µg/ml.
[b] Four days after rechallenge a pool of cells from three mice were tested. The results are expressed as mean of duplicate values and the difference between duplicate cultures are < 15% of mean. The IFN-γ release of cultures incubated without antigen was 390 pg/ml.
[c] A pool of CFP22 and CFP25 was tested.
[d] ND, not determined.

Skin test reaction in TB infected guinea pigs

The skin test activity of the purified proteins was tested in *M. tuberculosis* infected guinea pigs.

1 group of guinea pigs was infected via an ear vein with 1 x 10^4 CFU of *M. tuberculosis* H37Rv in 0,2 ml PBS. After 4

weeks skin tests were performed and 24 hours after injection erythema diameter was measured.

As seen in TABLES 6 and 6a all of the antigens induced a significant Delayed Type Hypersensitivity (DTH) reaction.

TABLE 6

DTH erythema diameter in guinea pigs infected with 1 x 10^4 CFU of *M. tuberculosis*, after stimulation with native antigens.

Antigen[a]	Skin reaction (mm)[b]
Control	2.00
PPD[c]	15.40 (0.53)
CFP7	ND[e]
CFP9	ND
CFP17	11.25 (0.84)
CFP20	8.88 (0.13)
CFP21	12.44 (0.79)
CFP22 + CFP25[d]	9.19 (3.10)
CFP26 (MPT51)	ND
CFP28	2.90 (1.28)
CFP29	6.63 (0.88)

The values presented are the mean of erythema diameter of four animals and the SEM's are indicated in the brackets. For PPD and CFP29 the values are mean of erythema diameter of ten animals.

[a] The antigens were tested in a concentration of 0,1 µg except for CFP29 which was tested in a concentration of 0,8 µg.

[b] The skin reactions are measured in mm erythema 24 h after intradermal injection.

[c] 10 TU of PPD was used.

[d] A pool of CFP22 and CFP25 was tested.

[e] ND, not determined.

Together these analyses indicate that most of the antigens identified were highly biologically active and recognized during TB infection in different animal models.

TABLE 6a

DTH erythema diameter of recombinant antigens in outbred guinea pigs infected with 1×10^4 CFU of *M. Tuberculosis*.

Antigen[a]	Skin reaction (mm)[b]	
Control	2.9	(0.3)
PPD[c]	14.5	(1.0)
CFP 7a	13.6	(1.4)
CFP 17	6.8	(1.9)
CFP 20	6.4	(1.4)
CFP 21	5.3	(0.7)
CFP 25	16.8	(0.8)
CFP 29	7.4	(2.2)
MPT 51	4.9	(1.1)

The values presented are the mean of erythema diameter of four animals and the SEM's are indicated in the brackets. For Control, PPD, and CFP 20 the values are mean of erythema diameter of eight animals.

[a] The antigens were tested in a concentration of 1,0 µg.

[b] The skin test reactions are measured in mm erythema 24 h after intradermal infection.

[c] 10 TU of PPD was used.

Biological activity of the purified recombinant antigens.

Interferon-γ induction in the mouse model of TB infection.

Primary infections. 8 to 12 weeks old female C57BL/6j(H-2^b), CBA/J(H-2^k), DBA.2(H-2^d) and A.SW(H-2^s) mice (Bomholtegaard, Ry) were given intravenous infections via the lateral tail vein with an inoculum of 5×10^4 *M. tuberculosis* suspended in PBS in a vol. of 0.1 ml. 14 days postinfection the animals were sacrificed and spleen cells were isolated and tested for the recognition of recombinant antigen.

As seen in TABLE 7 the recombinant antigens rCFP7A, rCFP17, rCFP21, rCFP25, and rCFP29 were all recognized in at least two strains of mice at a level comparable to ST-CF. rMPT51 and rCFP7 were only recognized in one or two strains respectively, at a level corresponding to no more than 1/3 of the

response detected after ST-CF stimulation. Neither of the antigens rCFP20 and rCFP22 were recognized by any of the four mouse strains.

Memory responses. 8-12 weeks old female C57BL/6j($H\text{-}2^b$) mice (Bomholtegaard, Ry) were given intravenous infections via the lateral tail vein with an inoculum of 5×10^4 *M. tuberculosis* suspended in PBS in a vol. of 0.1 ml. After 1 month of infection the mice were treated with isoniazid (Merck and Co., Rahway, NJ) and rifabutin (Farmatalia Carlo Erba, Milano, Italy) in the drinking water, for two months. The mice were rested for 4-6 months before being used in experiments. For the study of the recall of memory immunity, animals were infected with an inoculum of 1×10^6 bacteria i.v. and sacrificed at day 4 postinfection. Spleen cells were isolated and tested for the recognition of recombinant antigen.
As seen from TABLE 8, IFN-γ release after stimulation with rCFP17, rCFP21 and rCFP25 was at the same level as seen from spleen cells stimulated with ST-CF. Stimulation with rCFP7, rCFP7A and rCFP29 all resulted in an IFN-γ no higher than 1/3 of the response seen with ST-CF. rCFP22 was not recognized by IFN-γ producing cells. None of the antigens stimulated IFN-γ release in naive mice. Additionally non of the antigens were toxic to the cell cultures.

TABLE 7. T cell responses in primary TB infection.

Name	c57BL/6J($H2^b$)	DBA.2($H2^d$)	CBA/J($H2^k$)	A.SW($H2^s$)
rCFP7	+	+	-	-
rCFP7A	+++	+++	+++	+
rCFP17	+++	+	+++	+
rCFP20	-	-	-	-
rCFP21	+++	+++	+++	+
rCFP22	-	-	-	-
rCFP25	+++	++	+++	+
rCFP29	+++	+++	+++	++
rMPT51	+	-	-	-

Mouse IFN-γ release during recall of memory immunity to *M. tuberculosis*.

-:no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

TABLE 8. T cell responses in memory immune animals.

Name	Memory response
rCFP7	+
rCFP7A	++
rCFP17	+++
rCFP21	+++
rCFP22	-
rCFP29	+
rCFP25	+++
rMPT51	+

Mouse IFN-γ release 14 days after primary infection with *M. tuberculosis*.

-:no response; +: 1/3 of ST-CF; ++: 2/3 of ST-CF; +++: level of ST-CF.

Interferon-γ induction in human TB patients and BCG vaccinated people.

Human donors: PBMC were obtained from healthy BCG vaccinated donors with no known exposure to patients with TB and from patients with culture or microscopy proven infection with *Mycobacterium tuberculosis*. Blood samples were drawn from the TB patients 1-4 months after diagnosis.

Lymphocyte preparations and cell culture: PBMC were freshly isolated by gradient centrifugation of heparinized blood on Lymphoprep (Nycomed, Oslo, Norway). The cells were resuspended in complete medium: RPMI 1640 (Gibco, Grand Island, N.Y.) supplemented with 40 μg/ml streptomycin, 40 U/ml penicillin, and 0.04 mM/ml glutamine, (all from Gibco Laboratories, Paisley, Scotland) and 10% normal human ABO serum (NHS) from the local blood bank. The number and the viability of the cells were determined by trypan blue staining. Cultures were established with $2,5 \times 10^5$ PBMC in 200 μl in microtitre plates (Nunc, Roskilde, Denmark) and stimulated with no antigen, ST-CF, PPD (2.5μg/ml); rCFP7, rCFP7A, rCFP17, rCFP20, rCFP21, rCFP22, rCFP25, rCFP26, rCFP29, in a final concentration of 5 μg/ml. Phytohaemagglutinin, 1 μg/ml (PHA, Difco laboratories, Detroit, MI. was used as a positive control. Supernatants for the detection of cytokines were harvested after 5 days of culture, pooled and stored at -80°C until use.

Cytokine analysis: Interferon-γ (IFN-γ) was measured with a standard ELISA technique using a commercially available pair of mAb´s from Endogen and used according to the instructions for use. Recombinant IFN-γ (Gibco laboratories) was used as a standard. The detection level for the assay was 50 pg/ml. The variation between the duplicate wells did not exceed 10 % of the mean. Responses of 9 individual donors are shown in TABLE 9.

As seen in TABLE 9 high levels of IFN-γ release are obtained after stimulation with several of the recombinant antigens. rCFP7a and rCFP17 gives rise to responses comparable to STCF in almost all donors. rCFP7 seems to be most strongly recognized by BCG vaccinated healthy donors. rCFP21, rCFP25, rCFP26, and rCFP29 gives riss to a mixed picture with intermediate responses in each group, whereas low responses are obtained by rCFP20 and rCFP22.

TABLE 9. Mean values of results from the stimulation of human blood cells from 7 BCG vaccinated and 7 TB patients with recombinant antigens. SE values are given for each antigen. ST-CF and *M. avium* culture filtrate are shown for the comparison.

Controls, Healthy, BCG vaccinated, no known TB exposure

donor:	no ag	PHA	PPD	STCF	CFP7	CFP17	CFP7A	CFP20	CFP21	CFP22	CFP25	CFP26	CFP29
1	6	9564	8774	3966	7034	69	1793	58	152	73	182	946	86
2	48	12486	6603	8067	3146	10044	5267	29	6149	51	1937	526	2065
3	190	11929	10030	8299	8015	31863	8641	437	3194	569	2531	8076	6098
4	10	21029	4196	3537	1323	1939	5211	1	284	1	1344	20	125
5	1	18750	14209	13027	17725	8038	19002	1	3808	1	2103	974	8181

TB patients, 1-4 month after diagnosis

	no ag	PHA	PPD	STCF	CFP7	CFP17	CFP7A	CFP20	CFP21	CFP22	CFP25	CFP26	CFP29
6	9	8573	5096	6145	852	4250	4019	284	1131	48	2400	1078	4584
7	1	12413	6281	3393	168	6375	4505	11	4335	16	3082	1370	5115
8	4	11915	7671	7379	104	2753	3356	119	407	437	2069	712	5284
9	32	22130	16417	17213	8450	9783	16319	91	5957	67	10043	13313	9953

Example 6A

Four groups of 6-8 weeks old, female C57Bl/6J mice (Bomholtegård, Denmark) were immunized subcutaneously at the base of the tail with vaccines of the following compositions:

Group 1: 10 μg ESAT-6/DDA (250 μg)
Group 2: 10 μg MPT59/DDA (250μg)
Group 3: 10 μg MPT59-ESAT-6 /DDA (250 μg)
Group 4: Adjuvant control group: DDA (250 μg) in NaCl

The animals were injected with a volume of 0.2 ml. Two weeks after the first injection and 3 weeks after the second injection the mice were boosted a little further up the back. One week after the last immunization the mice were bled and the blood cells were isolated. The immune response induced was monitored by release of IFN-γ into the culture supernatants when stimulated in vitro with relevant antigens (see the following table).

Immunogen 10 μg/dose	For restimulation[a]: Ag in vitro			
	no antigen	ST-CF	ESAT-6	MPT59
ESAT-6	219 ± 219	569 ± 569	835 ± 633	-
MPT59	0	802 ± 182	-	5647 ± 159
Hybrid: MPT59-ESAT-6	127 ± 127	7453 ± 581	15133 ± 861	16363 ± 1902

[a] Blood cells were isolated 1 week after the last immunization and the release of IFN-γ (pg/ml) after 72h of antigen stimulation (5 μg/ml) was measured.
The values shown are mean of triplicates performed on cells pooled from three mice ± SEM

[b] - not determined

The experiment demonstrates that immunization with the hybrid stimulates T cells which recognize ESAT-6 and MPT59 stronger than after single antigen immunization. Especially the recognition of ESAT-6 was enhanced by immunization with the MPT59-ESAT-6 hybrid. IFN-γ release in control mice immunized with DDA never exceeded 1000 pg/ml.

EXAMPLE 6B

The recombinant antigens were tested individually as subunit vaccines in mice. Eleven groups of 6-8 weeks old, female C57Bl/6j mice (Bomholtegård, Denmark) were immunized subcutaneously at the base of the tail with vaccines of the following composition:

Group 1: 10 μg CFP7
Group 2: 10 μg CFP17
Group 3: 10 μg CFP21
Group 4: 10 μg CFP22
Group 5: 10 μg CFP25
Group 6: 10 μg CFP29
Group 7: 10 μg MPT51
Group 8: 50 μg ST-CF
Group 9: Adjuvant control group
Group 10: BCG 2,5 x 10^6/ml, 0,2 ml
Group 11: Control group: Untreated

All the subunit vaccines were given with DDA as adjuvant. The animals were vaccinated with a volume of 0.2 ml. Two weeks after the first injection and three weeks after the second injection group 1-9 were boosted a little further up the back. One week after the last injection the mice were bled and the blood cells were isolated. The immune response induced was monitored by release of IFN-γ into the culture supernatant when stimulated in vitro with the homologous protein.

6 weeks after the last immunization the mice were aerosol challenged with 5 x 10^6 viable *Mycobacterium tuberculosis*/ml. After 6 weeks of infection the mice were killed and the number of viable bacteria in lung and spleen of infected mice was determined by plating serial 3-fold dilutions of organ homogenates on 7H11 plates. Colonies were counted after 2-3 weeks of incubation. The protective efficacy is expressed as the difference between $\log_{10}$ values of the geometric mean of

counts obtained from five mice of the relevant group and the geometric mean of counts obtained from five mouse of the relevant control group.

The results from the experiments are presented in the following table.

Immunogenicity and protective efficacy in mice, of ST-CF and 7 subunit vaccines

Subunit Vaccine	Immunogenicity	Protective efficacy
ST-CF	+++	+++
CFP7	++	-
CFP17	+++	+++
CFP21	+++	++
CFP22	-	-
CFP25	+++	+++
CFP29	+++	+++
MPT51	+++	++

+++ Strong immunogen / high protection (level of BCG)
++ Medium immunogen / medium protection
- No recognition / no protection

In conclusion, we have identified a number of proteins inducing high levels of protection. Three of these CFP17, CFP25 and CFP29 giving rise to similar levels of protection as ST-CF and BCG while two proteins CFP21 and MPT51 induces protections around 2/3 the level of BCG and ST-CF. Two of the proteins CFP7 and CFP22 did not induce protection in the mouse model.

EXAMPLE 7

Species distribution of cfp7, cfp8, mpt51, rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf6, rd1-orf9a and rd1-orf9b as well as of cfp7a, cfp7b, cfp10a, cfp17, cfp20, cfp21, cfp22, cfp22a, cfp23, cfp25 and cfp25a.

Pregence of *cfp7*, *cfp9*, *mpt51*, *rd1-orf2*, *rd1-orf3*, *rd1-orf4*, *rd1-orf5*, *rd1-orf8*, *rd1-orf9a* and *rd1-orf9b* in different mycobacterial species.

In order to determine the distribution of the *cfp7*, *cfp9*, mpt51, *rd1-orf2*, *rd1-orf3*, *rd1-orf4*, *rd1-orf5*, *rd1-orf8*, *rd1-orf9a* and *rd1-orf9b* genes in species belonging to the *M. tuberculosis*-complex and in other mycobacteria PCR and/or Southern blotting was used. The bacterial strains used are listed in TABLE 10. Genomic DNA was prepared from mycobacterial cells as described previously (Andersen et al. 1992).

PCR analyses were used in order to determine the distribution of the *cfp7*, *cfp9* and *mpt51* gene in species belonging to the tuberculosis-complex and in other mycobacteria. The bacterial strains used are listed in TABLE 10. PCR was performed on genomic DNA prepared from mycobacterial cells as described previously (Andersen et al., 1992).

The oligonucleotide primers used were synthesised automatically on a DNA synthesizer (Applied Biosystems, Forster City, Ca, ABI-391, PCR-mode), deblocked, and purified by ethanol precipitation. The primers used for the analyses are shown in TABLE 11.

The PCR amplification was carried out in a thermal reactor (Rapid cycler, Idaho Technology, Idaho) by mixing 20 ng chromosomal with the mastermix (contained 0.5 μM of each oligonucleotide primer, 0.25 μM BSA (Stratagene), low salt buffer (20 mM Tris-HCl, pH8.8 , 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$ and 0,1% Triton X-100) (Stratagene), 0.25 mM of each deoxynucleoside triphosphate and 0.5 U Taq Plus Long DNA polymerase (Stratagene)). Final volume was 10 μl (all concentrations given are concentrations in the final volume). Predenaturation was carried out at 94°C for 30 s. 30 cycles of the following was performed: Denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min.

The following primer combinations were used (the length of the amplified products are given in parentheses):

mpt51: MPT51-3 and MPT51-2 (820 bp), MPT51-3 and MPT51-6 (108 bp), MPT51-5 and MPT51-4 (415 bp), MPT51-7 and MPT51-4 (325 bp).

cfp7: pVF1 and PVR1 (274 bp), pVF1 and PVR2 (197 bp), pVF3 and PVR1 (302 bp), pVF3 and PVR2 (125 bp).

cfp9: stR3 and stF1 (351 bp).

TABLE 10.
Mycobacterial strains used in this Example.

Species and strain(s)		Source
1. *M. tuberculosis*	H37Rv (ATCC 27294)	ATCC[a]
2.	H37Ra (ATCC 25177)	ATCC
3.	Erdman	Obtained from A. Lazlo, Ottawa, Canada
4. *M. bovis* BCG substrain; Danish 1331		SSI[b]
5.	Chinese	SSI[c]
6.	Canadian	SSI[c]
7.	Glaxo	SSI[c]
8.	Russia	SSI[c]
9.	Pasteur	SSI[c]
10.	Japan	WHO[c]
11. *M. bovis* MNC 27		SSI[c]
12. *M. africanum*		Isolated from a Danish patient
13. *M. leprae* (armadillo-derived)		Obtained from J. M. Colston, London, UK
14. *M. avium* (ATCC 15769)		ATCC
15. *M. kansasii* (ATCC 12478)		ATCC
16. *M. marinum* (ATCC 927)		ATCC
17. *M. scrofulaceum* (ATCC 19275)		ATCC
18. *M. intercellulare* (ATCC 15985)		ATCC
19. *M. fortuitum* (ATCC 6841)		ATCC
20. *M. xenopi*		Isolated from a Danish patient
21. *M. flavescens*		Isolated from a Danish patient
22. *M. szulgai*		Isolated from a Danish patient
23. *M. terrae*		SSI[c]
24. *E. coli*		SSI[d]
25. *S.aureus*		SSI[d]

[a] American Type Culture Collection, USA.
[b] Statens Serum Institut, Copenhagen, Denmark.

[c] Our collection Department of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark.
[d] Department of Clinical Microbiology, Statens Serum Institut, Denmark.
[e] WHO International Laboratory for Biological Standards, Statens Serum Institut, Copenhagen, Denmark.

TABLE 11.

Sequence of the *mpt51*, *cfp7* and *cfp9* oligonucleotides.

Orientation and oligonucleotide		Sequences (5'-3')[a]	Position[b] (nucleotides)
Sense			
	MPT51-1 (SEQ ID NO: 28)	CTCGAATTCGCCGGGTGCACACAG	6 - 21 (SEQ ID NO: 41)
	MPT51-3 (SEQ ID NO: 29)	CTCGAATTCGCCCCATACGAGAAC	143 - 158 (SEQ ID NO: 41)
	MPT51-5 (SEQ ID NO: 30)	GTGTATCTGCTGGAC	228 - 242 (SEQ ID NO: 41)
	MPT51-7 (SEQ ID NO: 31)	CCGACTGGCTGGCCG	418 - 432 (SEQ ID NO: 41)
	pvR1 (SEQ ID NO: 35)	GTACGAGAATTCATGTCGCAAATGATG	91 - 105 (SEQ ID NO: 1)
	pvR2 (SEQ ID NO: 36)	GTACGAGAATTCGAGCTTGGGGTGCCG	168 - 181 (SEQ ID NO: 1)
	stR3 (SEQ ID NO: 37)	CGATTCCAAGCTTGTGGCCGCCGACCCG	141 - 155 (SEQ ID NO: 3)
Antisense			
	MPT51-2 (SEQ ID NO: 32)	GAGGAATTCGCTTAGCGGATCGCA	946 - 932 (SEQ ID NO: 41)
	MPT51-4 (SEQ ID NO: 33)	CCCACATTCCGTTGG	642 - 628 (SEQ ID NO: 41)
	MPT51-6 (SEQ ID NO: 34)	GTCCAGCAGATACAC	242 - 228 (SEQ ID NO: 41)
	pvF1 (SEQ ID NO: 38)	CGTTAGGGATCCTCATCGCCATGGTGTTGG	340 - 323 (SEQ ID NO: 1)
	pvF3 (SEQ ID NO: 39)	CGTTAGGGATCCTGGTTCCACTGTGCC	268 - 255 (SEQ ID NO: 1)
	stF3 (SEQ ID NO: 40)	CGTTAGGGATCCTCAGGTCTTTTCGATG	467 - 452 (SEQ ID NO: 3)

[a] Nucleotides underlined are not contained in the nucleotide sequences of *mpt51*, *cfp7*, and *cfp9*.
[b] The positions referred to are of the non-underlined parts of the primers and correspond to the nucleotide sequence shown in SEQ ID NOs: 41, 1, and 3 for *mpt51*, *cfp7*, and *cfp9*, respectively.

The Southern blotting was carried out as described previously (Oettinger and Andersen, 1994) with the following modifications: 2 µg of genomic DNA was digested with *Pvu*II, electrophoresed in an 0.8% agarose gel, and transferred onto a nylon membrane (Hybond N-plus; Amersham International plc, Little Chalfont, United Kingdom) with a vacuum transfer device (Milliblot, TM-v; Millipore Corp., Bedford, MA). The *cfp7*,

cfp9, mpt51, *rdl-orf2*, *rdl-orf3*, *rdl-orf4*, *rdl-orf5*, *rdl-orf8*, *rdl-orf9a* and *rdl-orf9b* gene fragments were amplified by PCR from the plasmids pRVN01, pRVN02, pTO52, pTO87, pTO88, pT089, pT090, pT091, pT096 or pTO98 by using the primers shown in TABLE 11 and TABLE 2 (in Example 2a). The probes were labelled non-radioactively with an enhanced chemiluminescence kit (ECL; Amersham International plc, Little Chalfont, United Kingdom). Hybridization and detection was performed according to the instructions provided by the manufacturer. The results are summarized in TABLES 12 and 13.

TABLE 12. Interspecies analysis of the *cfp7*, *cfp9* and *mpt51* genes by PCR and/or Southern blotting and of MPT51 protein by Western blotting.

		PCR			Southern Blot			Western blot
	Species and strain	*cfp7*	*cfp9*	mpt51	*cfp7*	*cfp9*	mpt51	MPT51
1.	*M. tub.* H37Rv	+	+	+	+	+	+	+
2.	*M. tub.* H37Ra	+	+	+	N.D.	N.D.	+	+
3.	*M. tub.* Erdmann	+	+	+	+	+	+	+
4.	*M. bovis*	+	+	+			+	+
5.	*M. bovis* BCG Danish 1331	+	+	+	+	+	+	+
6.	*M. bovis* BCG Japan	+	+	N.D.	+	+	+	N.D.
7.	*M. bovis* BCG Chinese	+	+	N.D.	+	+	N.D.	N.D.
8.	*M. bovis* BCG Canadian	+	+	N.D.	+	+	N.D.	N.D.
9.	*M. bovis* BCG Glaxo	+	+	N.D.	+	+	N.D.	N.D.
10.	*M. bovis* BCG Russia	+	+	N.D.	+	+	N.D.	N.D.
11.	*M. bovis* BCG Pasteur	+	+	N.D.	+	+	N.D.	N.D.
12.	*M. africanum*	+	+	+	+	+	+	+
13.	*M. lepraemurium*	-	-	-	-	-	-	-
14.	*M. avium*	+	+	-	+	+	+	-
15.	*M. kansasii*	+	-	-	+	+	+	-
16.	*M. marinum*	-	(+)	-	+	+	+	-
17.	*M. scrofulaceum*	-	-	-	-	-	-	-
18.	*M. intracellulare*	+	(+)	-	+	+	+	-
19.	*M. fortuitum*	-	-	-	-	-	-	-
20.	*M. flavescens*	+	(+)	-	+	+	+	N.D.
21.	*M. xenopi*	-	-	-	N.D.	N.D.	+	-
22.	*M. szulgai*	(+)	(+)	-	-	+	-	-
23.	*M. terrae*	-	-	N.D.	N.D.	N.D.	N.D.	N.D.

+, positive reaction, -, no reaction, N.D. not determined.

cfp7, *cfp9* and *mpt51* were found in the *M. tuberculosis* complex including BCG and the environmental mycobacteria; *M. avium*, *M. kansasii*, *M. marinum*, *M. intracellular* and *M. flavescens*. *cfp9* was additionally found in *M. szulgai* and *mpt51* in *M. xenopi*.

Furthermore the presence of native MPT51 in culture filtrates from different mycobacterial strains was investigated with western blots developed with Mab HBT4.

There is a strong band at around 26 kDa in *M. tuberculosis* H37Rv, Ra, Erdman, *M. bovis* AN5, *M. bovis* BCG substrain Danish 1331 and *M. africanum*. No band was seen in the region in any other tested mycobacterial strains.

TABLE 13a. Interspecies analysis of the *rd1-orf2, rd1-orf3, rd1-orf4, rd1-orf5, rd1-orf8, rd1-orf9a* and *rd1-orf9b* genes by Southern blotting.

Species and strain	*rd1-orf2*	*rd1-orf3*	*rd1-orf4*	*rd1-orf5*	*rd1-orf8*	*rd1-orf9a*	*rd1-orf9b*
1. M. tub. H37Rv	+	+	+	+	+	+	+
2. M. bovis	+	+	+	+	N.D.	+	+
3. M. bovis BCG Danish 1331	+	-	-	-	N.D.	-	-
4. M. bovis BCG Japan	+	-	-	-	N.D.	-	-
5. M. avium	-	-	-	-	N.D.	-	-
6. M. kansasii	-	-	-	-	N.D.	-	-
7. M. marinum	+	-	+	-	N.D.	-	-
8. M. scrofulaceum	+	-	-	-	N.D.	-	-
9. M. intercellulare	-	-	-	-	N.D.	-	-
10. M. fortuitum	-	-	-	-	N.D.	-	-
11. M. xenopi	-	-	-	-	N.D.	-	-
12. M. szulgai	+	-	-	-	N.D.	-	-

+, positive reaction; -, no reaction, N.D. not determined.

Positive results for *rd1-orf2*, *rd1-orf3*, *rd1-orf4*, *rd1-orf5*, *rd1-orf8*, *rd1-orf9a* and *rd1-orf9b* were only obtained when using genomic DNA from *M. tuberculosis* and *M. bovis*, and not from *M. bovis* BCG or other mycobacteria analyzed except *rd1-orf4* which also was found in *M. marinum*.

Presence of *cfp7a*, *cfp7b*, *cfp10a*, *cfp17*, *cfp20*, *cfp21*, *cfp22*, *cfp22a*, *cfp23*, *cfp25* and *cfp25a* in different mycobacterial species.

Southern blotting was carried out as described for *rd1-orf2*, *rd1-orf3*, *rd1-orf4*, *rd1-orf5*, *rd1-orf8*, *rd1-orf9a* and *rd1-orf9b*. The *cfp7a*, *cfp7b*, *cfp10a*, *cfp17*, *cfp20*, *cfp21*, *cfp22*, *cfp22a*, *cfp23*, *cfp25* and *cfp25a* gene fragments were amplified by PCR from the recombinant pMCT6 plasmids encoding the individual genes. The primers used (same as the primers used for cloning) are described in example 3, 3A and 3B. The results are summarized in Table 13b.

TABLE 13b. Interspecies analysis of the *cfp7a*, *cfp7b*, *cfp10a*, *cfp17*, *cfp20*, *cfp21*, *cfp22*, *cfp22a*, *cfp23*, *cfp25*, and *cfp25a* genes by Southern blotting

Species and strain	*cfp7a*	*cfp7b*	*cfp10a*	*cfp17*	*cfp20*	*cfp21*	*cfp22*	*cfp22a*	*cfp23*	*cfp25*	*cfp25a*
1. M. tub. H37Rv	+	+	+	+	+	+	+	+	+	+	+
2. M. bovis	+	+	+	+	+	+	+	+	+	+	+
3. M. bovis BCG Danish 1331	+	+	+	+	+	N.D.	+	+	+	+	+
4. M. bovis BCG Japan	+	+	+	+	+	+	+	+	+	+	+
5. M. avium	+	N.D.	-	+	-	+	+	+	+	+	-
6. M. hominis	-	N.D.	+	-	-	-	+	-	+	-	-
7. M. marinum	+	+	-	-	+	+	+	+	+	+	+
8. M. scrofulaceum	-	-	+	-	+	+	-	+	+	+	-
9. M. intracellulare	+	+	-	+	-	+	+	-	+	+	-
10. M. fortuitum	-	N.D.	-	-	-	-	-	-	+	-	-
11. M. xenopi	+	+	+	+	+	+	+	-	+	+	+
12. M. szulgai	+	+	-	+	+	+	+	+	+	+	+

+, positive reaction; -, no reaction, N.D. not determined

LIST OF REFERENCES

Andersen, P. and Heron, I, 1993, J. Immunol. Methods **161**: 29-39.

Andersen, Å. B. *et al.*, 1992, Infect. Immun. **60**: 2317-2323.

Andersen P., 1994, Infect. Immun. **62**: 2536-44.

Andersen P. *et al.*, 1995, J. Immunol. **154**: 3359-72

Barkholt, V. and Jensen, A. L., 1989, Anal. Biochem. **177**: 318-322.

Borodovsky, M., and J. McIninch. 1993, Computers Chem. **17**: 123-133.

van Dyke M. W. *et al.*, 1992. Gene pp. 99-104.

Gosselin *et al.*, 1992, J. Immunol. **149**: 3477-3481.

Harboe, M. *et al.*, 1996, Infect. Immun. **64**: 16-22.

von Heijne, G., 1984, J. Mol. Biol. **173**: 243-251.

Hochstrasser, D.F. *et al.*, 1988, Anal.Biochem. **173**: 424-435

Köhler, G. and Milstein, C., 1975, Nature **256**: 495-497.

Li, H. *et al.*, 1993, Infect. Immun. **61**: 1730-1734.

Lindblad E.B. *et al.*, 1997, Infect. Immun. **65**: 623-629.

Mahairas, G. G. *et al.*, 1996, J. Bacteriol **178**: 1274-1282.

Maniatis T. *et al.*, 1989, "Molecular cloning: a laboratory manual", 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Nagai, S. *et al.*, 1991, Infect. Immun. **59**: 372-382.

Oettinger, T. and Andersen, Å. B., 1994, Infect. Immun. **62**: 2058-2064.

Ohara, N. *et al.*, 1995, Scand. J. immunol. **41**: 233-442.

Pal P. G. and Horwitz M. A., 1992, Infect. Immun. **60**: 4781-92.

Pearson, W. R. and Lipman D. J., 1988. Proc. Natl. Acad. Sci. USA **85**: 2444-2448.

Ploug, M. *et al.*, 1989, Anal. Biochem. **181**: 33-39.

Porath, J. *et al.*, 1985, FEBS Lett. **185**: 306-310.

Roberts, A.D. *et al.*, 1995, Immunol. **85**: 502-508.

Sørensen, A.L. *et al.*, 1995, Infect. Immun. **63**: 1710-1717.

Theisen, M. *et al.*, 1995, Clinical and Diagnostic Laboratory Immunology, **2**: 30-34.

Valdés-Stauber, N. and Scherer, S., 1994, Appl. Environ. Microbiol. **60**: 3809-3814.

Valdés-Stauber, N. and Scherer, S., 1996, Appl. Environ. Microbiol. **62**: 1283-1286.

Williams, N., 1996, Science **272**: 27.

Young, R. A. *et al.*, 1985, Proc. Natl. Acad. Sci. USA **82**: 2583-2587.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Statens Seruminstitut
(B) STREET: Artillerivej 5
(C) CITY: Copenhagen
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): 2300 S

(ii) TITLE OF INVENTION: Nucleic acid fragments and polypeptide fragments derived from M. tuberculosis

(iii) NUMBER OF SEQUENCES: 173

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(B) STRAIN: H37Rv

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 91..381

(ix) FEATURE:
(A) NAME/KEY: -35_signal
(B) LOCATION: 14..19

(ix) FEATURE:
(A) NAME/KEY: -10_signal
(B) LOCATION: 47..52

(ix) FEATURE:
(A) NAME/KEY: RBS
(B) LOCATION: 78..84

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 91..381

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GGCCGCCGGT ACCTATGTGG CCGCCGATGC TGCGGACGCG TCGACCTATA CCGGGTTCTG      60

ATCGAACCCT GCTGACCGAG AGGACTTGTG ATG TCG CAA ATC ATG TAC AAC TAC      114
                                 Met Ser Gln Ile Met Tyr Asn Tyr
                                   1               5

CCC GCG ATG TTG GGT CAC GCC GGG GAT ATG GCC GGA TAT GCC GGC ACG      162
Pro Ala Met Leu Gly His Ala Gly Asp Met Ala Gly Tyr Ala Gly Thr
     10                  15                  20

CTG CAG AGC TTG GGT GCC GAG ATC GCC GTG GAG CAG GCC GCG TTG CAG      210
Leu Gln Ser Leu Gly Ala Glu Ile Ala Val Glu Gln Ala Ala Leu Gln
 25                  30                  35                  40

AGT GCG TGG CAG GGC GAT ACC GGG ATC ACG TAT CAG GCG TGG CAG GCA      258
Ser Ala Trp Gln Gly Asp Thr Gly Ile Thr Tyr Gln Ala Trp Gln Ala
                 45                  50                  55

CAG TGG AAC CAG GCC ATG GAA GAT TTG GTG CGG GCC TAT CAT GCG ATG      306
Gln Trp Asn Gln Ala Met Glu Asp Leu Val Arg Ala Tyr His Ala Met
             60                  65                  70

TCC AGC ACC CAT GAA GCC AAC ACC ATG GCG ATG ATG GCC CGC GAC ACC      354
Ser Ser Thr His Glu Ala Asn Thr Met Ala Met Met Ala Arg Asp Thr
         75                  80                  85

GCC GAA GCC GCC AAA TGG GGC GGC TAG                                  381
Ala Glu Ala Ala Lys Trp Gly Gly
     90                  95
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ser Gln Ile Met Tyr Asn Tyr Pro Ala Met Leu Gly His Ala Gly
  1               5                  10                  15

Asp Met Ala Gly Tyr Ala Gly Thr Leu Gln Ser Leu Gly Ala Glu Ile
             20                  25                  30

Ala Val Glu Gln Ala Ala Leu Gln Ser Ala Trp Gln Gly Asp Thr Gly
         35                  40                  45

Ile Thr Tyr Gln Ala Trp Gln Ala Gln Trp Asn Gln Ala Met Glu Asp
     50                  55                  60

Leu Val Arg Ala Tyr His Ala Met Ser Ser Thr His Glu Ala Asn Thr
 65                  70                  75                  80
```

```
Met Ala Met Met Ala Arg Asp Thr Ala Glu Ala Ala Lys Trp Gly Gly
                85                  90                  95


(2) INFORMATION FOR SEQ ID NO: 3:

     (i) SEQUENCE CHARACTERISTICS:
           (A) LENGTH: 467 base pairs
           (B) TYPE: nucleic acid
           (C) STRANDEDNESS: double
           (D) TOPOLOGY: circular

    (ii) MOLECULE TYPE: DNA (genomic)

    (vi) ORIGINAL SOURCE:
           (A) ORGANISM: Mycobacterium tuberculosis
           (B) STRAIN: H37Rv

   (ix) FEATURE:
           (A) NAME/KEY: CDS
           (B) LOCATION: 141..467

   (ix) FEATURE:
           (A) NAME/KEY: -10_signal
           (B) LOCATION: 73..78

   (ix) FEATURE:
           (A) NAME/KEY: -35_signal
           (B) LOCATION: 4..9

   (ix) FEATURE:
           (A) NAME/KEY: RBS
           (B) LOCATION: 123..130

   (ix) FEATURE:
           (A) NAME/KEY: mat_peptide
           (B) LOCATION: 141..467


    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGGTAGCCGG ACCACGGCTG GGCAAAGATG TGCAGGCCGC CATCAAGGCG GTCAAGGCCG      60

GCGACGGCGT CATAAACCCG GACGGCACCT TGTTGGCGGG CCCCGCGGTG CTGACGCCCG     120

ACGAGTACAA CTCCCGGCTG GTG GCC GCC GAC CCG GAG TCC ACC GCG GCG          170
                    Met Ala Ala Asp Pro Glu Ser Thr Ala Ala
                      1               5                  10

TTG CCC GAC GGC GCC GGG CTG GTC GTT CTG GAT GGC ACC GTC ACT GCC       218
Leu Pro Asp Gly Ala Gly Leu Val Val Leu Asp Gly Thr Val Thr Ala
            15                  20                  25

GAA CTC GAA GCC GAG GGC TGG GCC AAA GAT CGC ATC CGC GAA CTG CAA       266
Glu Leu Glu Ala Glu Gly Trp Ala Lys Asp Arg Ile Arg Glu Leu Gln
        30                  35                  40
```

```
GAG CTG CGT AAG TCG ACC GGG CTG GAC GTT TCC GAC CGC ATC CGG GTG      314
Glu Leu Arg Lys Ser Thr Gly Leu Asp Val Ser Asp Arg Ile Arg Val
        45                  50                  55

GTG ATG TCG GTG CCT GCG GAA CGC GAA GAC TGG GCG CGC ACC CAT CGC      362
Val Met Ser Val Pro Ala Glu Arg Glu Asp Trp Ala Arg Thr His Arg
    60                  65                  70

GAC CTC ATT GCC GGA GAA ATC TTG GCT ACC GAC TTC GAA TTC GCC GAC      410
Asp Leu Ile Ala Gly Glu Ile Leu Ala Thr Asp Phe Glu Phe Ala Asp
 75                  80                  85                  90

CTC GCC GAT GGT GTG GCC ATC GGC GAC GGC GTG CGG GTA AGC ATC GAA      458
Leu Ala Asp Gly Val Ala Ile Gly Asp Gly Val Arg Val Ser Ile Glu
                95                 100                 105

AAG ACC TGA                                                          467
Lys Thr
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ala Ala Asp Pro Glu Ser Thr Ala Ala Leu Pro Asp Gly Ala Gly
  1               5                  10                  15

Leu Val Val Leu Asp Gly Thr Val Thr Ala Glu Leu Glu Ala Glu Gly
            20                  25                  30

Trp Ala Lys Asp Arg Ile Arg Glu Leu Gln Glu Leu Arg Lys Ser Thr
        35                  40                  45

Gly Leu Asp Val Ser Asp Arg Ile Arg Val Val Met Ser Val Pro Ala
    50                  55                  60

Glu Arg Glu Asp Trp Ala Arg Thr His Arg Asp Leu Ile Ala Gly Glu
65                  70                  75                  80

Ile Leu Ala Thr Asp Phe Glu Phe Ala Asp Leu Ala Asp Gly Val Ala
                85                  90                  95

Ile Gly Asp Gly Val Arg Val Ser Ile Glu Lys Thr
            100                 105
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(B) STRAIN: H37Rv

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 201..689

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 201..290

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 291..689

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
CGGGTCTGCA CGGATCCGGG CCGGGCAGGG CAATCGAGCC TGGGATCCGC TGGGGTGCGC      60

ACATCGCGGA CCCGTGCGCG GTACGGTCGA GACAGCGGCA CGAGAAAGTA GTAAGGGCGA     120

TAATAGGCGG TAAAGAGTAG CGGGAAGCCG GCCGAACGAC TCGGTCAGAC AACGCCACAG     180

CGGCCAGTGA GGAGCAGCGG GTG ACG GAC ATG AAC CCG GAT ATT GAG AAG          230
                      Met Thr Asp Met Asn Pro Asp Ile Glu Lys
                     -30                 -25

GAC CAG ACC TCC GAT GAA GTC ACG GTA GAG ACG ACC TCC GTC TTC CGC       278
Asp Gln Thr Ser Asp Glu Val Thr Val Glu Thr Thr Ser Val Phe Arg
-20                 -15                 -10                  -5

GCA GAC TTC CTC AGC GAG CTG GAC GCT CCT GCG CAA GCG GGT ACG GAG       326
Ala Asp Phe Leu Ser Glu Leu Asp Ala Pro Ala Gln Ala Gly Thr Glu
                 1               5                  10

AGC GCG GTC TCC GGG GTG GAA GGG CTC CCG CCG GGC TCG GCG TTG CTG       374
Ser Ala Val Ser Gly Val Glu Gly Leu Pro Pro Gly Ser Ala Leu Leu
        15                  20                  25

GTA GTC AAA CGA GGC CCC AAC GCC GGG TCC CGG TTC CTA CTC GAC CAA       422
Val Val Lys Arg Gly Pro Asn Ala Gly Ser Arg Phe Leu Leu Asp Gln
    30                  35                  40

GCC ATC ACG TCG GCT GGT CGG CAT CCC GAC AGC GAC ATA TTT CTC GAC       470
Ala Ile Thr Ser Ala Gly Arg His Pro Asp Ser Asp Ile Phe Leu Asp
 45                  50                  55                  60

GAC GTG ACC GTG AGC CGT CGC CAT GCT GAA TTC CGG TTG GAA AAC AAC       518
Asp Val Thr Val Ser Arg Arg His Ala Glu Phe Arg Leu Glu Asn Asn
                 65                  70                  75
```

```
GAA TTC AAT GTC GTC GAT GTC GGG AGT CTC AAC GGC ACC TAC GTC AAC      566
Glu Phe Asn Val Val Asp Val Gly Ser Leu Asn Gly Thr Tyr Val Asn
                80                  85                  90

CGC GAG CCC GTG GAT TCG GCG GTG CTG GCG AAC GGC GAC GAG GTC CAG     614
Arg Glu Pro Val Asp Ser Ala Val Leu Ala Asn Gly Asp Glu Val Gln
        95                  100                 105

ATC GGC AAG TTC CGG TTG GTG TTC TTG ACC GGA CCC AAG CAA GGC GAG     662
Ile Gly Lys Phe Arg Leu Val Phe Leu Thr Gly Pro Lys Gln Gly Glu
    110                 115                 120

GAT GAC GGG AGT ACC GGG GGC CCG TGA GCGCACCCGA TAGCCCCGCG           709
Asp Asp Gly Ser Thr Gly Gly Pro
125                 130

CTGGCCGGGA TGTCGATCGG GGCGGTCCTC GACCTGCTAC GACCGGATTT TCCTGATGTC   769

ACCATCTCCA AGATTCGATT CTTGGAGGCT GAGGGTCTGG TGACGCCCCG GCGGGCCTCA   829

TCGGGGTATC GGCGGTTCAC CGCATACGAC TGCGCACGGC TGCGATTCAT TCTCACTGCC   889
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Thr Asp Met Asn Pro Asp Ile Glu Lys Asp Gln Thr Ser Asp Glu
-30                 -25                 -20                 -15

Val Thr Val Glu Thr Thr Ser Val Phe Arg Ala Asp Phe Leu Ser Glu
                -10                 -5                  1

Leu Asp Ala Pro Ala Gln Ala Gly Thr Glu Ser Ala Val Ser Gly Val
        5                   10                  15

Glu Gly Leu Pro Pro Gly Ser Ala Leu Leu Val Val Lys Arg Gly Pro
    20                  25                  30

Asn Ala Gly Ser Arg Phe Leu Leu Asp Gln Ala Ile Thr Ser Ala Gly
35                  40                  45                  50

Arg His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val Ser Arg
                55                  60                  65

Arg His Ala Glu Phe Arg Leu Glu Asn Asn Glu Phe Asn Val Val Asp
            70                  75                  80

Val Gly Ser Leu Asn Gly Thr Tyr Val Asn Arg Glu Pro Val Asp Ser
        85                  90                  95
```

```
Ala Val Leu Ala Asn Gly Asp Glu Val Gln Ile Gly Lys Phe Arg Leu
    100                 105                 110

Val Phe Leu Thr Gly Pro Lys Gln Gly Glu Asp Asp Gly Ser Thr Gly
115                 120                 125                 130

Gly Pro
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 898 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(B) STRAIN: H37Rv

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 201..698

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 201..698

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
TCGACTCCGG CGCCACCGGG CAGGATCACG GTGTCGACGG GGTCGCCGGG GAATCCCACG      60

ATAACCACTC TTCGCGCCAT GAATGCCAGT GTTGGCCAAG CGCTGGCCTG GCGTCCACGC     120

CACACACCGC ACAGATTAGG ACACGCCGGC GCCGCAGCCC TGCCCGAAAG ACCGTGCACC     180

GGTCTTGGCA GACTGTGCCC ATG GCA CAG ATA ACC CTG CGA GGA AAC GCG          230
                      Met Ala Gln Ile Thr Leu Arg Gly Asn Ala
                        1               5                  10

ATC AAT ACC GTC GGT GAG CTA CCT GCT GTC GGA TCC CCG GCC CCG GCC       278
Ile Asn Thr Val Gly Glu Leu Pro Ala Val Gly Ser Pro Ala Pro Ala
                    15                  20                  25

TTC ACC CTG ACC GGG GGC GAT CTG GGG GTG ATC AGC AGC GAC CAG TTC       326
Phe Thr Leu Thr Gly Gly Asp Leu Gly Val Ile Ser Ser Asp Gln Phe
                30                  35                  40

CGG GGT AAG TCC GTG TTG CTG AAC ATC TTT CCA TCC GTG GAC ACA CCG       374
Arg Gly Lys Ser Val Leu Leu Asn Ile Phe Pro Ser Val Asp Thr Pro
        45                  50                  55
```

```
GTG TGC GCG ACG AGT GTG CGA ACC TTC GAC GAG CGT GCG GCG GCA AGT       422
Val Cys Ala Thr Ser Val Arg Thr Phe Asp Glu Arg Ala Ala Ala Ser
     60                  65                  70

GGC GCT ACC GTG CTG TGT GTC TCG AAG GAT CTG CCG TTC GCC CAG AAG       470
Gly Ala Thr Val Leu Cys Val Ser Lys Asp Leu Pro Phe Ala Gln Lys
 75                  80                  85                  90

CGC TTC TGC GGC GCC GAG GGC ACC GAA AAC GTC ATG CCC GCG TCG GCA       518
Arg Phe Cys Gly Ala Glu Gly Thr Glu Asn Val Met Pro Ala Ser Ala
                 95                 100                 105

TTC CGG GAC AGC TTC GGC CAG GAT TAC GGC GTG ACC ATC GCC GAC GGG       566
Phe Arg Asp Ser Phe Gly Glu Asp Tyr Gly Val Thr Ile Ala Asp Gly
            110                 115                 120

CCG ATG GCC GGG CTG CTC GCC CGC GCA ATC GTG GTG ATC GGC GCG GAC       614
Pro Met Ala Gly Leu Leu Ala Arg Ala Ile Val Val Ile Gly Ala Asp
        125                 130                 135

GGC AAC GTC GCC TAC ACG GAA TTG GTG CCG GAA ATC GCG CAA GAA CCC       662
Gly Asn Val Ala Tyr Thr Glu Leu Val Pro Glu Ile Ala Gln Glu Pro
    140                 145                 150

AAC TAC GAA GCG GCG CTG GCC GCG CTG GGC GCC TAG GCTTTCACAA            708
Asn Tyr Glu Ala Ala Leu Ala Ala Leu Gly Ala
155                 160                 165

GCCCCGCGCG TTCGGCGAGC AGCGCACGAT TTCGAGCGCT GCTCCCGAAA AGCGCCTCGG     768

TGGTCTTGGC CCGCCGGTAA TACAGGTGCA GGTCGTGCTC CCACGTGAAG GCGATGGCAC     828

CGTGGATCTG AAGAGCGGAG CCGGCGCATA ACACAAAGGT TTCCGCGGTC TGCGCCTTCG     888

CCAGCGGCGC                                                            898
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 165 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Ala Gln Ile Thr Leu Arg Gly Asn Ala Ile Asn Thr Val Gly Glu
  1               5                  10                  15

Leu Pro Ala Val Gly Ser Pro Ala Pro Ala Phe Thr Leu Thr Gly Gly
             20                  25                  30

Asp Leu Gly Val Ile Ser Ser Asp Gln Phe Arg Gly Lys Ser Val Leu
         35                  40                  45

Leu Asn Ile Phe Pro Ser Val Asp Thr Pro Val Cys Ala Thr Ser Val
     50                  55                  60
```

```
Arg Thr Phe Asp Glu Arg Ala Ala Ala Ser Gly Ala Thr Val Leu Cys
 65                  70                  75                  80

Val Ser Lys Asp Leu Pro Phe Ala Gln Lys Arg Phe Cys Gly Ala Glu
                85                  90                  95

Gly Thr Glu Asn Val Met Pro Ala Ser Ala Phe Arg Asp Ser Phe Gly
            100                 105                 110

Glu Asp Tyr Gly Val Thr Ile Ala Asp Gly Pro Met Ala Gly Leu Leu
        115                 120                 125

Ala Arg Ala Ile Val Val Ile Gly Ala Asp Gly Asn Val Ala Tyr Thr
    130                 135                 140

Glu Leu Val Pro Glu Ile Ala Gln Glu Pro Asn Tyr Glu Ala Ala Leu
145                 150                 155                 160

Ala Ala Leu Gly Ala
                165
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(B) STRAIN: H37Rv

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 201..854

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 201..296

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 297..854

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
ATAATCAGCT CACCGTTGGG ACCGACCTCG ACCAGGGGTC CTTTGTGACT GCCGGGCTTG      60

ACGCGGACGA CCACAGAGTC GGTCATCGCC TAAGGCTACC GTTCTGACCT GGGGCTGCGT     120

GGGCGCCGAC GACGTGAGGC ACGTCATGTC TCAGCGGCCC ACCGCCACCT CGGTCGCCGG     180
```

```
CAGTATGTCA GCATGTGCAG ATG ACT CCA CGC AGC CTT GTT CGC ATC GTT          230
                      Met Thr Pro Arg Ser Leu Val Arg Ile Val
                     -32     -30             -25

GGT GTC GTG GTT GCG ACG ACC TTG GCG CTG GTG AGC GCA CCC GCC GGC        278
Gly Val Val Val Ala Thr Thr Leu Ala Leu Val Ser Ala Pro Ala Gly
        -20             -15             -10

GGT CGT GCC GCG CAT GCG GAT CCG TGT TCG GAC ATC GCG GTC GTT TTC        326
Gly Arg Ala Ala His Ala Asp Pro Cys Ser Asp Ile Ala Val Val Phe
    -5              1               5                   10

GCT CGC GGC ACG CAT CAG GCT TCT GGT CTT GGC GAC GTC GGT GAG GCG        374
Ala Arg Gly Thr His Gln Ala Ser Gly Leu Gly Asp Val Gly Glu Ala
            15                  20                  25

TTC GTC GAC TCG CTT ACC TCG CAA GTT GGC GGG CGG TCG ATT GGG GTC        422
Phe Val Asp Ser Leu Thr Ser Gln Val Gly Gly Arg Ser Ile Gly Val
        30                  35                  40

TAC GCG GTG AAC TAC CCA GCA AGC GAC GAC TAC CGC GCG AGC GCG TCA        470
Tyr Ala Val Asn Tyr Pro Ala Ser Asp Asp Tyr Arg Ala Ser Ala Ser
    45                  50                  55

AAC GGT TCC GAT GAT GCG AGC GCC CAC ATC CAG CGC ACC GTC GCC AGC        518
Asn Gly Ser Asp Asp Ala Ser Ala His Ile Gln Arg Thr Val Ala Ser
60                  65                  70

TGC CCG AAC ACC AGG ATT GTG CTT GGT GGC TAT TCG CAG GGT GCC ACG        566
Cys Pro Asn Thr Arg Ile Val Leu Gly Gly Tyr Ser Gln Gly Ala Thr
 75                  80                  85                  90

GTC ATC GAT TTG TCC ACC TCG GCG ATG CCG CCC GCG GTG GCA GAT CAT        614
Val Ile Asp Leu Ser Thr Ser Ala Met Pro Pro Ala Val Ala Asp His
                95                  100                 105

GTC GCC GCT GTC GCC CTT TTC GGC GAG CCA TCC AGT GGT TTC TCC AGC        662
Val Ala Ala Val Ala Leu Phe Gly Glu Pro Ser Ser Gly Phe Ser Ser
            110                 115                 120

ATG TTG TGG GGC GGC GGG TCG TTG CCG ACA ATC GGT CCG CTG TAT AGC        710
Met Leu Trp Gly Gly Gly Ser Leu Pro Thr Ile Gly Pro Leu Tyr Ser
        125                 130                 135

TCT AAG ACC ATA AAC TTG TGT GCT CCC GAC GAT CCA ATA TGC ACC GGA        758
Ser Lys Thr Ile Asn Leu Cys Ala Pro Asp Asp Pro Ile Cys Thr Gly
    140                 145                 150

GGC GGC AAT ATT ATG GCG CAT GTT TCG TAT GTT CAG TCG GGG ATG ACA        806
Gly Gly Asn Ile Met Ala His Val Ser Tyr Val Gln Ser Gly Met Thr
155                 160                 165                 170

AGC CAG GCG GCG ACA TTC GCG GCG AAC AGG CTC GAT CAC GCC GGA TGA        854
Ser Gln Ala Ala Thr Phe Ala Ala Asn Arg Leu Asp His Ala Gly
                175                 180                 185

TCAAAGACTG TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA TGTACACCGG CTGGAATCTG      914
```

AAGGGCAAGA ACCCGGTATT CATCAGGCCG GATGAAATGA CGGTCGGGCG GTAATCGTTT 974

GTGTTGAACG CGTAGAGCCG ATCACCGCCG GGGCTGGTGT AGACCTCAAT GTTTGTGTTC 1034

GCCGGCAGGG TTCCGGATCC 1054

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Met Thr Pro Arg Ser Leu Val Arg Ile Val Gly Val Val Val Ala Thr
-32  -30                 -25                 -20

Thr Leu Ala Leu Val Ser Ala Pro Ala Gly Gly Arg Ala Ala His Ala
  -15                 -10                  -5

Asp Pro Cys Ser Asp Ile Ala Val Val Phe Ala Arg Gly Thr His Gln
  1               5                  10                  15

Ala Ser Gly Leu Gly Asp Val Gly Glu Ala Phe Val Asp Ser Leu Thr
            20                  25                  30

Ser Gln Val Gly Gly Arg Ser Ile Gly Val Tyr Ala Val Asn Tyr Pro
        35                  40                  45

Ala Ser Asp Asp Tyr Arg Ala Ser Ala Ser Asn Gly Ser Asp Asp Ala
    50                  55                  60

Ser Ala His Ile Gln Arg Thr Val Ala Ser Cys Pro Asn Thr Arg Ile
65                  70                  75                  80

Val Leu Gly Gly Tyr Ser Gln Gly Ala Thr Val Ile Asp Leu Ser Thr
                85                  90                  95

Ser Ala Met Pro Pro Ala Val Ala Asp His Val Ala Ala Val Ala Leu
            100                 105                 110

Phe Gly Glu Pro Ser Ser Gly Phe Ser Ser Met Leu Trp Gly Gly Gly
        115                 120                 125

Ser Leu Pro Thr Ile Gly Pro Leu Tyr Ser Ser Lys Thr Ile Asn Leu
    130                 135                 140

Cys Ala Pro Asp Asp Pro Ile Cys Thr Gly Gly Gly Asn Ile Met Ala
145                 150                 155                 160

His Val Ser Tyr Val Gln Ser Gly Met Thr Ser Gln Ala Ala Thr Phe
                165                 170                 175

Ala Ala Asn Arg Leu Asp His Ala Gly
            180                 185
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 949 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(B) STRAIN: H37Rv

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 201..749

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 224..749

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
AGCCGCTCGC GTGGGGTCAA CCGGGTTTCC ACCTGCTCAC TCATTTTGCC GCCTTTCTGT      60

GTCCGGGCCG AGGCTTGCGC TCAATAACTC GGTCAAGTTC CTTCACAGAC TGCCATCACT     120

GGCCCGTCGG CGGGCTCGTT GCGGGTGCGC CGCGTGCGGG TTTGTGTTCC GGGCACCGGG     180

TGGGGGCCCG CCCGGGCGTA ATG GCA GAC TGT GAT TCC GTG ACT AAC AGC          230
                      Met Ala Asp Cys Asp Ser Val Thr Asn Ser
                       -7      -5                   1

CCC CTT GCG ACC GCT ACC GCC ACG CTG CAC ACT AAC CGC GGC GAC ATC       278
Pro Leu Ala Thr Ala Thr Ala Thr Leu His Thr Asn Arg Gly Asp Ile
      5                  10                  15

AAG ATC GCC CTG TTC GGA AAC CAT GCG CCC AAG ACC GTC GCC AAT TTT       326
Lys Ile Ala Leu Phe Gly Asn His Ala Pro Lys Thr Val Ala Asn Phe
 20                  25                  30                  35

GTG GGC CTT GCG CAG GGC ACC AAG GAC TAT TCG ACC CAA AAC GCA TCA       374
Val Gly Leu Ala Gln Gly Thr Lys Asp Tyr Ser Thr Gln Asn Ala Ser
                 40                  45                  50

GGT GGC CCG TCC GGC CCG TTC TAC GAC GGC GCG GTC TTT CAC CGG GTG       422
Gly Gly Pro Ser Gly Pro Phe Tyr Asp Gly Ala Val Phe His Arg Val
             55                  60                  65

ATC CAG GGC TTC ATG ATC CAG GGT GGC GAT CCA ACC GGG ACG GGT CGC       470
Ile Gln Gly Phe Met Ile Gln Gly Gly Asp Pro Thr Gly Thr Gly Arg
         70                  75                  80

GGC GGA CCC GGC TAC AAG TTC GCC GAC GAG TTC CAC CCC GAG CTG CAA       518
Gly Gly Pro Gly Tyr Lys Phe Ala Asp Glu Phe His Pro Glu Leu Gln
     85                  90                  95
```